Correspondence

Confirmation of Organized Modularity in the Yeast Interactome

Nicolas Bertin[☯], Nicolas Simonis[☯], Denis Dupuy[☯], Michael E. Cusick, Jing-Dong J. Han, Hunter B. Fraser[*], Frederick P. Roth[*], Marc Vidal[*]

A recent *PLoS Biology* article [1] rejected the conclusions of two previous publications [2,3] that two categories of highly connected "hub" proteins—"date" and "party" hubs—have distinct properties in the *Saccharomyces cerevisiae* interactome network. Currently available protein–protein interaction datasets are vastly incomplete, even for yeast [4]. Therefore, it is reasonable to rigorously re-scrutinize global properties of interactome networks as new datasets become available. Here we show that distinctions between date and party hubs [2], previously shown in a high-quality filtered yeast interactome (FYI) dataset [2,3], are in fact confirmed in an updated literature-curated yeast interactome network.

Data Quality

Two protein–protein interaction datasets were used in [1]: a high-confidence (HC) network obtained from both curated literature and high-throughput sources, and a subgraph of HC that was obtained by linking the nodes of FYI with HC edges (HC^{fyi}). As explained in [2], it is crucial that high-quality data be used to partition date and party hubs. Therefore, FYI was originally generated as the union of two high-confidence interaction datasets: one curated from small-scale studies published in the literature [5] and another obtained by stringently requiring support from at least two out of four sources of high-throughput interaction evidence [2]. We use a similar definition here to derive a filtered high-confidence ('filtered-HC') dataset containing 2,561 proteins linked by 5,996 interactions (Table S1) from HC. To eliminate false positive interactions that were either reported once but never confirmed or that were obtained through curation error, our analysis included literature-curated interactions only if they were observed in two independent articles (i.e., associated with two or more independent PubMed IDs). Moreover, many interactions in HC were derived from a single experiment reported in multiple publications—e.g., reference [6] describes an approximate superset of the experiments including those reported in reference [7]. Such publications [6–10] were considered dependent and merged. Thus 2,423 protein pairs were removed from HC. Also, we did not include interactions supported solely by high-throughput yeast two-hybrid screening [11,12] (97 pairs) or supported solely by high-throughput pull-down followed by mass spectrometry screening (742 pairs) [6–10,13] (see Table S1 for a complete list of interactions in filtered-HC).

Consistency of Date and Party Hub Classification across Datasets

We identified date and party hubs in both HC and filtered-HC (all analyses were also performed on the HC^{fyi} network; see Figure S1). Since both networks contain many new interactions relative to FYI, and since some erroneous interactions might have been corrected, the proteins originally identified as hubs in FYI cannot and should not be assumed to be identical. For the analyses described here, we therefore defined hubs anew using a degree threshold that includes the top 20% most connected nodes [2]. This corresponds to a degree of 10 or more for HC (19.4% of the proteins) and a degree of 7 or more for filtered-HC (21.7%).

In the original report of the date/party hub distinction [2], bimodality was observed in the average Pearson's correlation coefficient (AvgPCC) distribution of hubs for two out of five expression datasets examined [2]. The complete lack of bimodality observed in [1] may stem from a conservative statistical test that assumed a uniform unimodal null distribution. We emphasize that bimodality was not deemed essential evidence of the party/date hub distinction in the initial report [2].

Since party and date hubs fall along a continuum, the choice of an AvgPCC threshold that distinguishes them is somewhat arbitrary (although our previous conclusions were robust to this choice [2]). Therefore, we adopted the PCC threshold of 0.5 for all networks considered here (this is the same threshold applied previously to PCC distributions that did not appear bimodal [1,2]). Thirteen expression datasets [14–31] were considered in addition to the original five independent datasets [2] (see Table S2). Strikingly, 86% of the FYI-defined hubs found in filtered-HC retained their date/party designation (Figure 1A) (81% for HC). This indicates that assignment to one or the other category is robust across datasets.

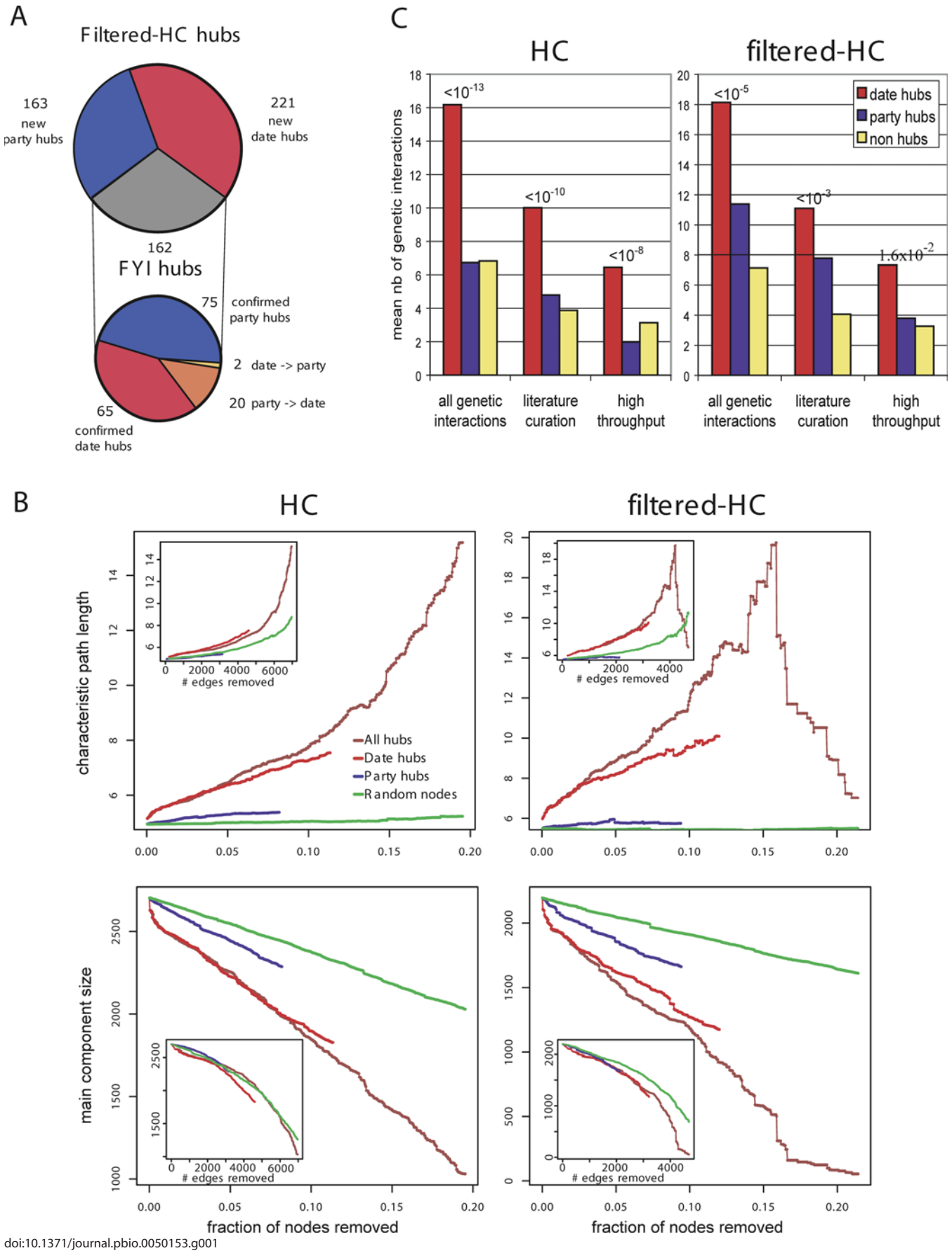
We suggest that some analyses presented in [1] (in particular the network tolerance to hub deletion) erred by not taking into account new hubs defined by the increased number of interactions relative to the original FYI. This strategy ignores 46% of the hubs in HC^{fyi} [1] and thus effectively immunizes them in the attack resistance analysis and eliminates them from the genetic interaction comparison.

Figure 1. Analysis of Hub Properties in the Filtered-HC and HC Interactome Networks

(A) Consistency of the party/date attribution between FYI and filtered-HC. Because filtered-HC network has many more interactions than FYI, only 162 of the 546 hubs in filtered-HC had been previously found in FYI. Filtered-HC confirmed 86% of the party/date designations in FYI. In addition, 20% of FYI hubs are not considered as hubs anymore in the new filtered-HC network because of the higher connectivity threshold.

(B) The effect on the characteristic path length (top panels) and main component size (bottom panels) of the networks upon gradual node removal for HC (left panels) and filtered-HC (right panels). Attacks against all hubs (brown curve), party hubs (blue curve), date hubs (red curve), and random nodes (green curve). Insets show an additional control for connectivity differences between categories with the x-axis representing the number of edges removed from the network.

(C) Date hubs participate in more genetic interactions than party hubs or non-hubs [2], as measured here by mean number of interactions [1] from a network of curated genetic interactions [32] for both filtered-HC (right panel) and HC (left panel). Inside each panel, bars show the number of genetic interactions held by date hubs (red), party hubs (blue), and non-hub proteins (yellow). The *p*-values assessing the difference of the means between date and party hubs (Mann-Whitney *U*-test) are indicated above the bars.



doi:10.1371/journal.pbio.0050153.g001

Distinct Topological Properties of Date and Party Hubs

When removed from the network, party and date hubs have strikingly distinct effects on the overall topology of HC, filtered-HC, and HC^{fyi}. Removing date hubs dramatically disrupts the characteristic path length (CPL) of the network, whereas removing party hubs has a negligible effect (Figure 1B), as previously observed [2]. Importantly, this difference in behavior is not sensitive to the specific threshold values of degree k and AvgPCC chosen here to define hubs and party hubs, respectively (Figure S2). The CPL of a network measures the mutual closeness of nodes in a network. The claim in [1] that date and party hub removal has an indistinguishable effect on network topology was based on the analysis of a different topological feature altogether—main component size. This is a poor measure of network clustering in that it does not, for example, discriminate an extended beads on a string topology from a completely connected clique. This measure is also highly sensitive to a single spurious interaction that connects two otherwise disconnected subgraphs. By contrast, the dramatic decrease in CPL that we observe for date hubs in HC, filtered-HC, and HC^{fyi} suggests their coordinating role and confirms the original findings [2].

Genetic Interactions

In [2] we showed that date hubs exhibit a higher genetic interaction density than party hubs. Reference [1] described analysis of two sets of genetic interactions: one from a union of high-throughput studies (HTP-GI), and another from the literature (LC-GI) [32]. Both LC-GI and HTP-GI datasets are potentially subject to bias since gene pairs were selected nonrandomly for testing, but these are the best datasets currently available. While the LC-GI analysis confirmed our original finding, the HTP-GI analysis did not [1], which we confirmed using date/party hubs defined from FYI. However, examining HTP-GI in the larger HC and filtered-HC networks, we find that date hubs in both HC and filtered-HC exhibit higher genetic interaction density than party hubs or non-hubs (Figure 1C), confirming the original report [2]. This difference remains after controlling for connectivity of hubs in the protein interaction network (Figure S3).

Evolutionary Rate

We also confirmed the difference in evolutionary rates [33] between date and party hubs that was reported previously [3]. Using the filtered-HC network (with hubs defined as above) we found that date hubs evolve significantly faster than party hubs (Wilcoxon $p = 0.01$). Furthermore, using our expanded expression dataset, the PCC of hubs was negatively correlated with their evolutionary rates (Pearson $r = -0.22$, $p = 1 \times 10^{-7}$), even when controlling for protein abundance [34] in either rich (Pearson partial $r = -0.19$, $p = 3 \times 10^{-6}$) or minimal media (Pearson partial $r = -0.20$, $p = 2 \times 10^{-6}$). The same result was obtained when considering the HC and HC^{fyi} networks (unpublished data). Moreover, a recent report independently supported evolutionary rate differences between date and party hub and explained these differences in terms of three-dimensional protein structure [35].

Summary

We confirmed that date and party hubs have different topological properties, with the coordinating role of date hubs being supported by a greater impact on CPL. We also confirmed that date hubs participate in more genetic interactions and evolve more rapidly than party hubs. These observations, as well as the identity of the nodes considered as date and party, remained largely consistent within all tested networks (HC, filtered-HC, HC^{fyi}), demonstrating the robustness of the results originally observed in [2]. Thus, this updated analysis confirms the validity of the distinction between date and party hubs in the yeast interactome [2,3], and shows that the date and party hub concept and the "stratus-like" network [1] model are not mutually exclusive. ■

Acknowledgements

Funding. This work was supported by the Keck Foundation (FPR and MV) and by NIH grants HG0017115 (FPR and MV) and HG003224 (FPR).

Competing interests. The authors have declared that no competing interests exist.

Supporting Information

Figure S1. Hub Deletion and Genetic Interaction Analysis for the HC^{fyi} Interaction Network as Defined in [1]

Found at doi:10.1371/journal.pbio.0050153.sg001 (172 KB PDF).

Figure S2. Different Effect on Gradual Date or Party Node Removal on the CPLs of the Networks for Filtered-HC Is Not Dependent on the PCC Threshold Used to Define Party Hubs.

Found at doi:10.1371/journal.pbio.0050153.sg002 (378 KB PDF).

Figure S3. Genetic Connectivity of Date and Party Hubs

(A) Mean number of genetic interactions reported corrected by the physical connectivity. (B) The mean absolute connectivity for each hub category and the genetic interaction connectivity normalized by the number of protein–protein interactions observed for all three protein–protein interaction datasets using either HTP-GI or LC-GI separately or combined. p-values assessing the difference of the means (Mann-Whitney U-test) are indicated.

Found at doi:10.1371/journal.pbio.0050153.sg003 (102 KB PDF).

Table S1. Filtered-HC Protein-Protein Interaction Dataset.

Found at doi:10.1371/journal.pbio.0050153.st001 (8.5 MB XLS).

Table S2. Filtered-HC Date and Party Hubs Degrees, Clustering Coefficients and AvgPCC Values for Each Microarray Dataset.

Found at doi:10.1371/journal.pbio.0050153.st002 (252 KB XLS).

References

1. Batada NN, Reguly T, Breitkreutz A, Boucher L, Breitkreutz BJ, et al. (2006) Stratus not altocumulus: A new view of the yeast protein interaction network. PLoS Biol 4(10): e317. doi:10.1371/journal.pbio.0040317.
2. Han JD, Bertin N, Hao T, Goldberg DS, Berriz GF, et al. (2004) Evidence for dynamically organized modularity in the yeast protein-protein interaction network. Nature 430: 88–93.
3. Fraser HB, Hirsh AE, Steinmetz LM, Scharfe C, Feldman MW (2002) Evolutionary rate in the protein interaction network. Science 296: 750–752.
4. Han JD, Dupuy D, Bertin N, Cusick ME, Vidal M (2005) Effect of sampling on topology predictions of protein-protein interaction networks. Nat Biotechnol 23: 839–844.
5. Mewes HW, Frishman D, Guldener U, Mannhaupt G, Mayer K, et al. (2002) MIPS: a database for genomes and protein sequences. Nucleic Acids Res 30: 31–34.
6. Gavin AC, Aloy P, Grandi P, Krause R, Boesche M, et al. (2006) Proteome survey reveals modularity of the yeast cell machinery. Nature 440: 631–636.
7. Gavin AC, Bosche M, Krause R, Grandi P, Marzioch M, et al. (2002) Functional organization of the yeast proteome by systematic analysis of protein complexes. Nature 415: 141–147.
8. Grandi P, Rybin V, Bassler J, Petfalski E, Strauss D, et al. (2002) 90S pre-ribosomes include the 35S pre-rRNA, the U3 snoRNP, and 40S subunit processing factors but predominantly lack 60S synthesis factors. Mol Cell 10: 105–115.
9. Krogan NJ, Peng WT, Cagney G, Robinson MD, Haw R, et al. (2004) High-definition macromolecular composition of yeast RNA-processing complexes. Mol Cell 13: 225–239.
10. Krogan NJ, Cagney G, Yu H, Zhong G, Guo X, et al. (2006) Global landscape of protein complexes in the yeast *Saccharomyces cerevisiae*. Nature 440: 637–643.

11. Ito T, Chiba T, Ozawa R, Yoshida M, Hattori M, et al. (2001) A comprehensive two-hybrid analysis to explore the yeast protein interactome. Proc Natl Acad Sci U S A 98: 4569–4574.
12. Uetz P, Giot L, Cagney G, Mansfield TA, Judson RS, et al. (2000) A comprehensive analysis of protein-protein interactions in *Saccharomyces cerevisiae*. Nature 403: 623–627.
13. Ho Y, Gruhler A, Heilbut A, Bader GD, Moore L, et al. (2002) Systematic identification of protein complexes in *Saccharomyces cerevisiae* by mass spectrometry. Nature 415: 180–183.
14. Gasch AP, Spellman PT, Kao CM, Carmel-Harel O, Eisen MB, et al. (2000) Genomic expression programs in the response of yeast cells to environmental changes. Mol Biol Cell 11: 4241–4257.
15. Spellman PT, Sherlock G, Zhang MQ, Iyer VR, Anders K, et al. (1998) Comprehensive identification of cell cycle-regulated genes of the yeast *Saccharomyces cerevisiae* by microarray hybridization. Mol Biol Cell 9: 3273–3297.
16. Roberts CJ, Nelson B, Marton MJ, Stoughton R, Meyer MR, et al. (2000) Signaling and circuitry of multiple MAPK pathways revealed by a matrix of global gene expression profiles. Science 287: 873–880.
17. Travers KJ, Patil CK, Wodicka L, Lockhart DJ, Weissman JS, et al. (2000) Functional and genomic analyses reveal an essential coordination between the unfolded protein response and ER-associated degradation. Cell 101: 249–258.
18. Chu S, DeRisi J, Eisen M, Mulholland J, Botstein D, et al. (1998) The transcriptional program of sporulation in budding yeast. Science 282: 699–705.
19. Chitikila C, Huisinga KL, Irvin JD, Basehoar AD, Pugh BF (2002) Interplay of TBP inhibitors in global transcriptional control. Mol Cell 10: 871–882.
20. Gasch AP, Huang M, Metzner S, Botstein D, Elledge SJ, et al. (2001) Genomic expression responses to DNA-damaging agents and the regulatory role of the yeast ATR homolog Mec1p. Mol Biol Cell 12: 2987–3003.
21. Mnaimneh S, Davierwala AP, Haynes J, Moffat J, Peng WT, et al. (2004) Exploration of essential gene functions via titratable promoter alleles. Cell 118: 31–44.
22. Segal E, Shapira M, Regev A, Pe'er D, Botstein D, et al. (2003) Module networks: Identifying regulatory modules and their condition-specific regulators from gene expression data. Nat Genet 34: 166–176.
23. Yoshimoto H, Saltsman K, Gasch AP, Li HX, Ogawa N, et al. (2002) Genome-wide analysis of gene expression regulated by the calcineurin/Crz1p signaling pathway in *Saccharomyces cerevisiae*. J Biol Chem 277: 31079–31088.
24. Haugen AC, Kelley R, Collins JB, Tucker CJ, Deng C, et al. (2004) Integrating phenotypic and expression profiles to map arsenic-response networks. Genome Biol 5: R95.
25. Lai LC, Kosorukoff AL, Burke PV, Kwast KE (2006) Metabolic-state-dependent remodeling of the transcriptome in response to anoxia and subsequent reoxygenation in *Saccharomyces cerevisiae*. Eukaryot Cell 5: 1468–1489.
26. Roberts GG, Hudson AP (2006) Transcriptome profiling of *Saccharomyces cerevisiae* during a transition from fermentative to glycerol-based respiratory growth reveals extensive metabolic and structural remodeling. Mol Genet Genomics 276: 170–186.
27. Wyrick JJ, Holstege FC, Jennings EG, Causton HC, Shore D, et al. (1999) Chromosomal landscape of nucleosome-dependent gene expression and silencing in yeast. Nature 402: 418–421.
28. Lee SE, Pellicioli A, Demeter J, Vaze MP, Gasch AP, et al. (2000) Arrest, adaptation, and recovery following a chromosome double-strand break in *Saccharomyces cerevisiae*. Cold Spring Harb Symp Quant Biol 65: 303–314.
29. Smith JJ, Marelli M, Christmas RH, Vizeacoumar FJ, Dilworth DJ, et al. (2002) Transcriptome profiling to identify genes involved in peroxisome assembly and function. J Cell Biol 158: 259–271.
30. Ogawa N, DeRisi J, Brown PO (2000) New components of a system for phosphate accumulation and polyphosphate metabolism in *Saccharomyces cerevisiae* revealed by genomic expression analysis. Mol Biol Cell 11: 4309–4321.
31. Huang J, Zhu H, Haggarty SJ, Spring DR, Hwang H, et al. (2004) Finding new components of the target of rapamycin (TOR) signaling network through chemical genetics and proteome chips. Proc Natl Acad Sci U S A 101: 16594–16599.
32. Reguly T, Breitkreutz A, Boucher L, Breitkreutz BJ, Hon GC, et al. (2006) Comprehensive curation and analysis of global interaction networks in *Saccharomyces cerevisiae*. J Biol 5: 11.
33. Hirsh AE, Fraser HB, Wall DP (2005) Adjusting for selection on synonymous sites in estimates of evolutionary distance. Mol Biol Evol 22: 174–177.
34. Newman JR, Ghaemmaghami S, Ihmels J, Breslow DK, Noble M, et al. (2006) Single-cell proteomic analysis of *S. cerevisiae* reveals the architecture of biological noise. Nature 441: 840–846.
35. Kim PM, Lu LJ, Xia Y, Gerstein MB (2006) Relating three-dimensional structures to protein networks provides evolutionary insights. Science 314: 1938–1941.

Citation: Bertin N, Simonis N, Dupuy D, Cusick ME, Han JDJ, et al. (2007) Confirmation of organized modularity in the yeast interactome. PLoS Biol 5(6): e153. doi:10.1371/journal.pbio.0050153

Nicolas Bertin, Nicolas Simonis, Denis Dupuy, Michael E. Cusick, and Marc Vidal are with the Center for Cancer Systems Biology (CCSB) and Department of Cancer Biology, Dana-Farber Cancer Institute, and Department of Genetics, Harvard Medical School, Boston, Massachusetts, United States of America. Jing-Dong J. Han is with the Chinese Academy of Sciences Key Laboratory for Molecular Developmental Biology, Institute of Genetics and Developmental Biology, Chinese Academy of Sciences, Beijing, China. Hunter B. Fraser is with the Broad Institute of Harvard and Massachusetts Institute of Technology, Cambridge, Massachusetts, United States of America. Frederick P. Roth is with the Department of Biological Chemistry and Molecular Pharmacology, Harvard Medical School and Center for Cancer Systems Biology (CCSB), Dana-Farber Cancer Institute, Boston, Massachusetts, United States of America.

☯ These authors contributed equally to this work.

* To whom correspondence should be addressed. E-mail: marc_vidal@dfci.harvard.edu (MV); hunter@broad.mit.edu (HBF); fritz_roth@hms.harvard.edu (FPR)

Still Stratus Not Altocumulus: Further Evidence against the Date/Party Hub Distinction

Nizar N. Batada*, Teresa Reguly, Ashton Breitkreutz, Lorrie Boucher, Bobby-Joe Breitkreutz, Laurence D. Hurst*, Mike Tyers*

Using a small dataset of protein–protein interactions [1], it was proposed that the yeast protein interaction network is made up of two sorts of hubs, party and date, and that these define modularity in the yeast protein interaction network. We found [2], by using several larger high-confidence datasets and appropriate statistical analyses, that we could not support these conclusions. Bertin et al. [3] now invite analysis of a further dataset of protein-protein interactions, which they argue support the party/date distinction. The claimed properties of party and date hubs are not, however, present in this dataset either. In particular, when controlling for important covariables where necessary, there is no evidence for (1) bimodality in partner co-expression, (2) enrichment for similarly localized proteins that physically interact with party hubs, (3) a lower rate of evolution of party hubs, (4) differences in the effects of deletion of date and party hubs, or (5) higher genetic connectivity of date hubs. In sum, all of our prior conclusions remain robust and there is no evidence for distinctive classes of network hubs.

It was suggested [1] that some hub proteins operate at the same intracellular place and time with their multiple interactants (as if at a party) while others operate on a one-by-one basis with their numerous partners (as if on a date). Is this distinction informative? Originally, four features were used to the draw a partition between date and party hubs: expression bimodality, localization entropy, network fragmentation, and genetic connectivity [1]. A subsequent analysis suggested a fifth distinction, namely different rates of evolution after control for covariables [4]. Given the small size of the original dataset and the absence of statistical support for some of the assertions, we asked [2] whether these claims were robust. In both the original dataset and in new high-confidence interaction datasets [2], we found we could not support any of the five points of evidence. Bertin et al. now nominate a new dataset, which they argue supports three of the five points of evidence.

Table 1. Test for Bimodality of Neighbour Correlation Distribution at Two Different Hub Connectivity Thresholds

Source of Expression Data	Dip Score (top 21.7%)	Significance	Dip Score (top 10%)	Significance
hughes00	0.009	ns	0.015	ns
mnaimneh04	0.008	ns	0.020	ns
gasch00	0.009	ns	0.017	ns
orourke04	0.008	ns	0.014	ns
spellman98	0.011	ns	0.012	ns
roberts00	0.009	ns	0.014	ns
gasch01	0.011	ns	0.013	ns
causton01	0.008	ns	0.013	ns
segal03	0.011	ns	0.013	ns
fleming02	0.015	ns	0.015	ns
schuller03	0.012	ns	0.015	ns
jellinski00	0.011	ns	0.019	ns
zhu00	0.010	ns	0.017	ns
primig00	0.009	ns	0.015	ns
yoshimoto02	0.008	ns	0.015	ns
ideker01	0.012	ns	0.018	ns
chitikila02	0.010	ns	0.023	ns
nautiyal02	0.009	ns	0.015	ns
shamji00	0.008	ns	0.016	ns
mutka01	0.009	ns	0.015	ns
mccammon03	0.011	ns	0.019	ns
chu98	0.010	ns	0.016	ns
gross00	0.009	ns	0.015	ns
ferea99	0.014	ns	0.018	ns
travers00	0.011	ns	0.018	ns

Result of dip tests for deviation from unimodality, defining hubs as the 21.7% most connected proteins in the dataset of Bertin et al ($n = 556$; column 2 and 3; dip score at the 95% level is 0.023 and 0.027 for 99%) and as the 10% most connected proteins in the dataset of Bertin et al ($n = 265$; column 4 and 5; dip score at the 95% level is 0.032 and 0.038 for 99%). References for expression datasets are listed in Supplementary Table 1. Dip test was as implemented in R [12].
ns, not significant.
doi:10.1371/journal.pbio.0050154.t001

Bertin et al. first note a curation issue with one of our many datasets, called HC, which inadvertently contains interactions that were, owing to an ambiguity in the literature, supported by a single analysis. We certainly agree that inclusion of the data from [5] and [6] as independent validations was in error, as the data in [6] indeed fully encompasses that of [5] (A-C Gavin, personal communication). However, approximately half of the interactions reported in [5] remain multi-validated by other means. An updated high-confidence dataset that removes this duplication and incorporates more recent interaction data is available here (Dataset S1) and as a download from the BioGRID database (see http://www.thebiogrid.org). Importantly, however, our dataset HC^m is unaffected by the above concern as we required validation of an interaction by multiple different methods. The new build of Bertin et al. (called "filtered-HC") mimics HC^m by excluding interactions not multivalidated with different methods. As the results of HC^m confirmed those of our other datasets [2], we were surprised by the claim that the date/party distinction is still supported in filtered-HC. Because this dataset provides the most robustly defendable set of interactions, here we re-analyse the filtered-HC network to ask whether it substantiates the date/party distinction.

No Evidence for Bimodality of Co-Expression Values

Han et al. originally proposed that clear evidence for a binary hub classification (party versus date) derives from bimodal distribution of co-expression (PCC) values: one class with high average PCCs (party hubs) and the other with low average PCCs (date hubs) [1]. This proposal was based exclusively on visual inspection of the data. By contrast, we applied a formal test that examines deviation from a null of unimodality [7,8] and found no evidence for bimodality. Up to now we have analysed 25 expression datasets across seven protein interaction builds (including filtered-HC; Table 1) and added datasets nominated by Han et al., a total of 181 separate tests. To a first approximation, by chance we should expect to see around nine incidences of significance at the 5% significance level owing to type I error (although this assumes independence between datasets). We find just two.

Given this lack of evidence for bimodality, Bertin et al. appear to concur that bimodality cannot be used to define party and date hubs. Surprisingly though, they now assert that bimodality never was a key point of evidence. The original definition of date and party hubs, however, stated that bimodality represented a "natural boundary" between the two classes [1]; indeed it was argued that the lack of obvious bimodality in some expression datasets was due to low sample sizes [1]. At the same time, Bertin et al. also venture to suggest that the standard statistical test for deviation from unimodality [7,8] has a high false negative rate. It does not (see Text S1).

Table 2. Opposite Localization Entropy of Date and Party Hubs

Dataset	L-,N-	L-,N+	L+,N-	L+,N+
date 5	1.3	0.34	0.94	0.37
party 5	1.3	0.27	0.89	0.37
p	0.08	0.01	0.86	0.77
date 25	1.2	0.33	0.68	0.29
party 25	1.3	0.36	1	0.4
p	0.13	0.19	0.001	0.01

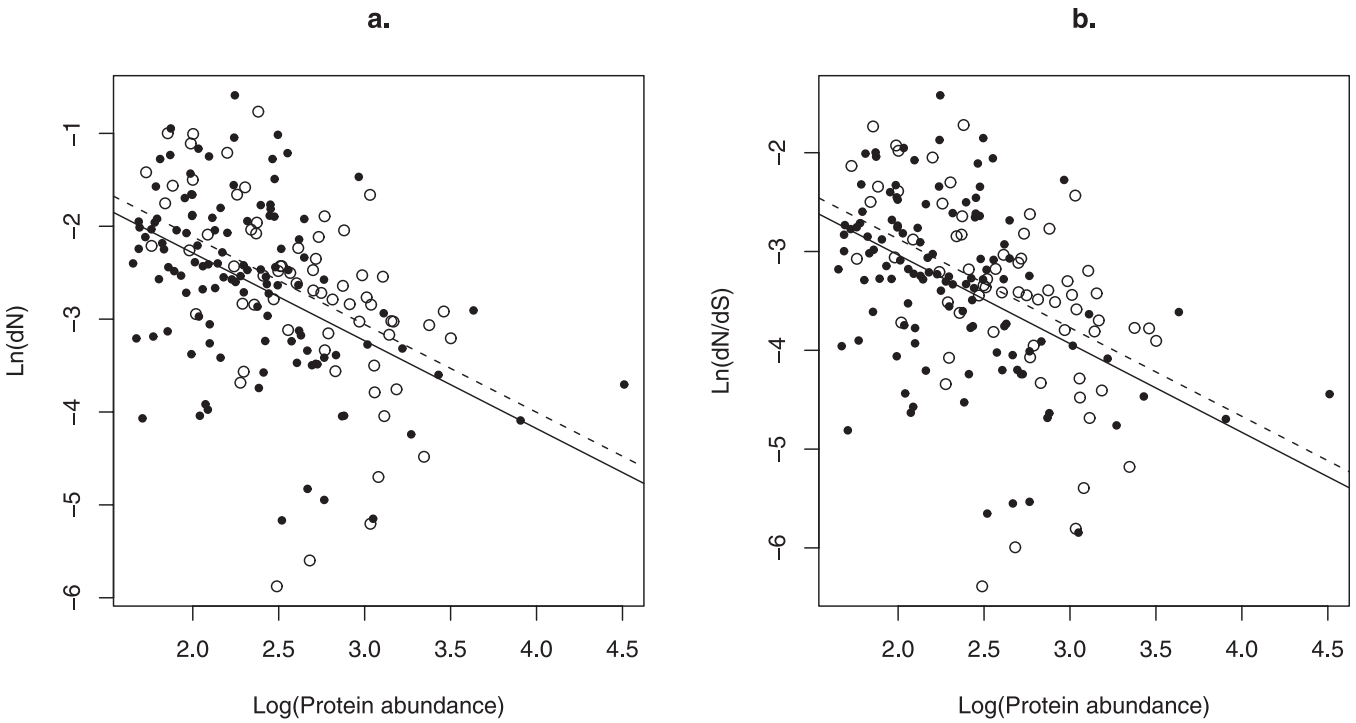
Two sets of data to define date and party hubs were used: the five expression datasets from Bertin et al., and 25 expression datasets employed here. L stands for localization and N stands for normalization; "+" indicates inclusion and "-" indicates exclusion. L+ means that cytoplasm and nucleus localizations were included; N+ means that normalization was done. L- and N- corresponds to the original Han et al. method, in which some localization data was omitted and data was not normalized [1]. Han et al reported higher entropy for date versus party hubs in the FYI dataset.
doi:10.1371/journal.pbio.0050154.t002

Neighbours of Date Hubs Do Not Have more Diverse Localizations

Originally, Han et al. reported that the partners of date hubs have more diverse intracellular localizations, as measured by information entropy [1]. However, this analysis did not normalise for connection density and arbitrarily omitted data from some cellular compartments [1]. In the filtered-HC dataset again (Table 2), as before [2], upon normalization and inclusion of all the data, the entropy is in the opposite direction to that predicted by the date/party hypothesis. This inversion we showed [2] is owing to differences in abundance that follow from the assignment of party hubs as those with highly co-expressed partners (PCC > 0.5). As Bertin et al. make no statement on this issue, we assume that they do not dispute this result.

Definitions and Inferences

The evidence for the biological relevance of date and party hubs falls into two classes: the definitional, namely bimodality/co-expression and subcellular colocalization, and



doi:10.1371/journal.pbio.0050154.g001

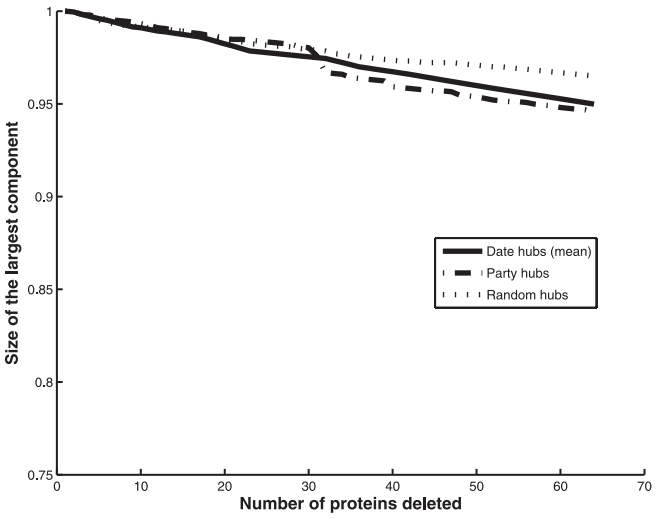
Figure 1. No Difference in the Evolutionary Rate of Date Hubs and Party Hubs

ANCOVA of natural log of rate of evolution, measured either as (A) *dN* or (B) *dN/dS* predicted by date/party distinction, with protein abundance as the covariate. The black line is for date, the dotted for party hubs. For (A), ANCOVA ln(*K*a) versus party/date with log10(abundance) as a covariate: effect of covariate, $t = 6.9$, $p \sim 8 \times 10{-}11$, effect of date/party, $t = 1.27$, $p = 0.21$. For (B), ANCOVA ln(*K*a/*K*s) versus party/date with log10(abundance) as a covariate: effect of covariate, $t = 6.99$, $p \sim 5 \times 10{-}11$, effect of date/party, t=1.24, $p = 0.21$. Note that taking the log of the variables on the y-axis forces loss of two data points (one party, one date) with $dN = 0$. However, results are unaffected by using, for example, ln(0.1 + *dN*) and ln(0.1 + *dN/dS*), which permits their inclusion. Similarly the residuals from the fit of *x* versus *y* are no different for date and party hubs for ln(0.1 + *dN*) residuals of date are if anything lower than those for party (−0.026 versus 0.46 but not significantly so, $p = 0.16$, *t*-test; ln(0.1 + *dN/dS*) mean for date is −0.015, for party 0.027, $p = 0.18$). We have repeated the analysis using a different outgroup *(S. bayanus)*, and still find no effect on covariate controlled analysis (unpublished data).

the inferential, or corollary behaviours that may derive from the underlying biology. As the definitional aspects do not bear scrutiny, one must be suspicious that any correlates are merely consequences of the method used to define the two hub classes. The only standing criterion left is the arbitrary distinction between those hubs with a PCC > 0.5 and those without. Highly co-expressed proteins do have a number of odd properties, namely higher connectivity and abundance. These biases are robust in the filtered HC dataset: party hubs have higher connectivity ($p = 0.00006$) and protein abundance ($p = 0.001$). It is then important to ask whether further properties stem from such biases.

Given their Abundance, Date and Party Hubs Do Not Evolve at Different Rates

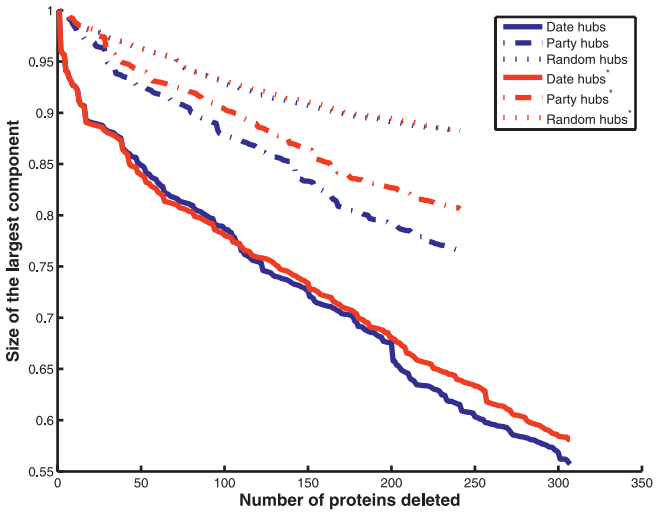
Bertin et al. find that party hubs evolve more slowly. As originally noted [4], the question is whether party hubs evolve slower than date hubs when controlling for important covariates, most notably protein abundance [9]. We showed previously that any weak tendency for party hubs to evolve slower was accounted for by their abundance [2]. Unlike our prior analysis, Bertin et al. do not ask if party and date hubs evolve at different rates controlling for abundance but, instead, ask if PCC is related to evolutionary rate controlling for abundance. However, they inappropriately apply a parametric test (Pearson product-moment correlation) that requires the distribution of all variables to be normally distributed.



doi:10.1371/journal.pbio.0050154.g002

Figure 2. Effect of Hub Deletion Controlling for Connectivity via Hub Deletion for Nonessential Party and Date Hubs

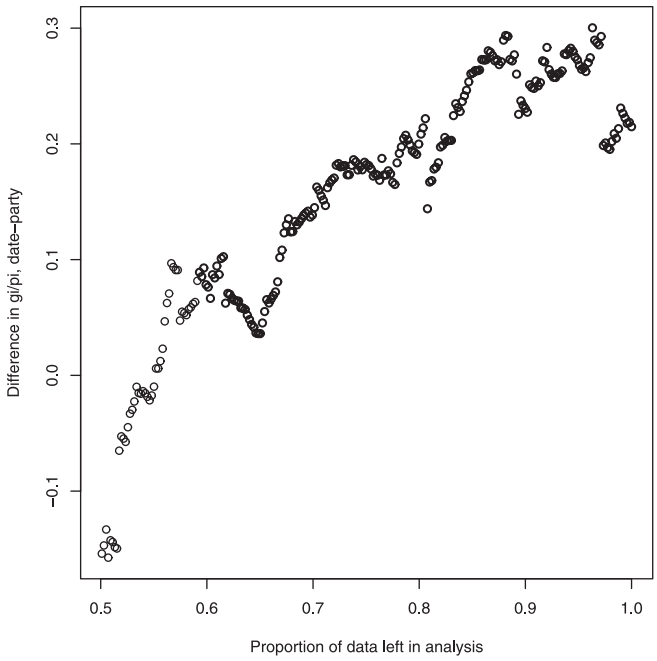
Date and party hub deletion effect on the integrity of the interaction network as measured by the relative size of the largest connected component (MCS) after deletion. Hubs were deleted in descending order by connectivity. Because the number of date hubs was much larger than number of party hubs (189 versus 64 respectively), we sampled the same number of date hubs as party hubs 200 times and determined the deletion effect each time. The mean effect of deletion of date hubs is plotted.



doi:10.1371/journal.pbio.0050154.g003

Figure 3. Effect of Hub Deletion Controlling for Connectivity via Hub Deletion before and after Random Swap of Date and Party Hubs of Similar Connectivity

Lines in red are after 50% swap of hubs, in blue for the original case. Because hub swapping has no effect, connectivity (not position in the network) explains why date and party have apparently different effects upon deletion.



doi:10.1371/journal.pbio.0050154.g004

Figure 4. The Effect of Study Bias on the Difference between the Mean Number of Genetic Interactions per Physical Connection (g_i/p_i) of Party and Date Hubs

We define bias as the difference between the number of independent validations of a genetic interaction of a given protein and the actual, nonredundant genetic connectivity, normalised by the nonredundant genetic connectivity [15]. We rank ordered all genes according to their study bias. We then eliminated the most biased data point and recalculated the difference in g_i/p_i for date versus party hubs for the remaining genes (reported on the *y*-axis). We then removed the next most biased, and so forth. At 0.5 residual, half of the original 489 genes were left in the analysis. Purging of the most biased genes removes any tendency for party and date hubs to differ; any possible difference between party and date hubs is hence owing to study bias.

Although the method is robust to some degree of deviation from normality, the extent to which the abundance data is non-normal is extreme (Shaprio-Wilks tests for null of normality, $W = 0.2$, $p \ll 0.0001$, $W = 1$ implying normality, $W \ll 1$ implying deviation from normality). This leaves two avenues: either to transform the data to make them approximately normal or to perform the equivalent non-parametric test.

Partial Spearman's correlation is the nonparametric equivalent. Using evolutionary rate data from *sensu strictu* yeasts [10], controlling for abundance [11], the more highly co-expressed genes have, if anything, a slightly higher rate of evolution (partial rho controlling for abundance, $rho = +0.13$, $p = 0.02$, p determined by simulation, implemented in R [11]). If we log transform the abundance data then the parametric correlation agrees that the sign of the partial correlation changes ($rho = +0.029$, $p = 0.36$). The log transformed abundance data has a Shapiro Wilks W score of 0.95, as opposed to 0.2 for the untransformed.

Our previous tests differed from that performed by Bertin et al: we employed analysis of covariance (ANCOVA) to ask whether date and party hubs evolve at different rates when covariate controlled (this being the prior claim [4]). We find, in accord with our results [2], that differences in abundance explain all difference in rates of evolution between the two classes (Fig 1). In the ANCOVA, as above, if anything date hubs evolve slightly slower than party hubs (Fig 1). Analysis of residuals supports these results (Fig 1). Although we can recover the result of Bertin et al. when Pearson's partial correlation is inappropriately applied to nontransformed data, all appropriate tests reject the contention of evolutionary rate differences.

Bertin et al. also suggest that a recent study of hub proteins that bind partners at multiple different sites, as opposed to re-use of the same site, provides support for the difference in evolutionary rate between party and date proteins [13]. However, this report failed to properly control for abundance [13], which if performed reveals no differences ($p > 0.45$) (Text S2). These results accord with our prior finding that, controlling for abundance, more highly connected hubs do not evolve more slowly, in no small part owing to re-use of binding sites [14]. In summary, evolutionary rate differences provide do not support the date/party distinction.

No Evidence for Large Differences in Effects of Hub Deletion when Allowing for Connectivity

It is argued that date hubs establish network integrity because of their positioning as intermodule linkers, as opposed to the intramodule positioning of party hubs [1]. But might any differences in deletion of date versus party hubs merely reflect a difference in connectivity of the two hub classes? Two metrics were used to measure the effect of hub deletion on the network: characteristic pathway length (CPL) and main component size (MCS) [1]. We previously considered [2] CPL to be of limited worth, because differences in pathway length may not have biological consequences (for example, since diffusion is fast, transmission delays due to increase in number of intermediate steps may be inconsequential). Moreover, CPL is susceptible to network incompleteness, which is acute for small stringent datasets such as filtered-HC. However, to enable comparison with Bertin et al., we analyze both MCS and CPL.

In addition to connectivity, it is desirable to correct for dispensability, because it is not biologically sensible to analyze networks that are fatally crippled by the loss of essential genes. Fortunately, nonessential date and party hubs have equal connectivity ($p = 0.94$), and thus control for both parameters simultaneously. Deletion of nonessential date and party hubs has an identical effect on network integrity (for MCS, see Figure 2; for CPL see Figure S1). Bertin et al. observe the same result for MCS even without controlling for dispensability. As an alternative means to correct for connectivity, we randomly swapped date and party hubs of the same connectivity. If the differential deletion effect is solely due to inter- versus intramodule positioning, then interchanging date with party hubs should obviate the difference. Instead, hub swapping yields the same deletion profile as the original unswapped case (Figure 3). Finally, we asked whether, even in the absence of controls for connectivity or dispensability, the difference between party and date hubs is sensitive to removal of just a few extreme hubs. Removal of just the top two percent of hubs obviates the difference between date and party hub deletion on MCS (Figure S2).

In sum, controlling for connectivity by two different means eliminates the difference between date and party hub deletion; even when not controlling for connectivity, the deletion effect relies entirely on a few extreme date hubs. There is thus no reason to suppose date and party hubs have different network positions.

No Evidence for a Difference in Genetic Connectivity

While Bertin et al. contend that date hubs have more genetic interactions in filtered-HC, they acknowledge that study bias may confound analysis, as noted [1]. Using a metric of study bias [15] (see Figure 4), we find that date and party hubs do indeed differ in their study bias ($p = 0.039$, Mann Whitney *U*-test). To examine the impact of this, we considered the difference in mean number of genetic interactions per physical connection (g_i/p_i) between date and party hubs; this metric controls for the fact that genetic and physical interactions are positively correlated [16]. As we incrementally purge the data of study bias, the difference in mean g_i/p_i between date and party hubs diminishes to zero (Figure 4). Even making no allowance for study bias, g_i/p_i for date and party hubs is not significantly different (Mann Whitney *U*-test; $p > 0.06$). There is thus no significant difference in genetic connectivity of party and date hubs.

Conservation of Date/Party Classification Is a Consequence of Definition, Not of Biology

Finally, Bertin et al. raise one new prospective line of evidence, namely, those proteins that appear as hubs across datasets tend to preserve their status as party or date. This observation, however, follows definition: if a hub is co-expressed (with PCC > 0.5) in any one dataset, it is defined as a party hub; if not, by default it is a date hub. Once a hub is classified as a party hub, its status cannot change solely with the addition of extra expression data. The reverse classification, i.e., date to party, is also disfavored because co-expression across different assays is not independent. The low rates of transfer of hub status merely follow from definitions and do not address the biological validity of the date/party distinction.

Summary

In the new filtered-HC dataset, as in others, the two definitional criteria of date/party hubs find no support. Four corollary points of evidence— rate of evolution, effect of deletion on network topology, genetic connectivity, and hub status quo—also find no support. That across multiple datasets, and under multiple different tests, we repeatedly find no evidence for the date/party hypothesis suggests that network hubs do not fall into discrete classes. ■

Supporting Information

Figure S1. Deletion of Nonessential Date and Party Hubs Does Not Have a Differential Effect on CPL in the Filtered-HC Dataset

Each indicated hub class was serially deleted in order of decreasing connectivity (percent of network deleted, *x*-axis) and CPL calculated after each deletion (*y*-axis) Deletion of either all hubs, date hubs, party hubs, or random nodes recapitulates previously reported effects on CPL. However, as for MCS analysis described in the main text, when connectivity is controlled for by deletion of only nonessential hubs, there is no differential contribution of date versus party hubs to CPL.

Found at doi:10.1371/journal.pbio.0050154.sg001 (1.1 MB EPS).

Figure S2. Effect of Hub Deletion before and after Removal of the Top 2% most Highly Connected Proteins in the Filtered-HC Network

Lines in blue are those prior to removal of top 2%, in red after removal. Note that the difference between date and party is very sensitive to the presence of very few extremely highly connected proteins, most of which are classified as date hubs.

Found at doi:10.1371/journal.pbio.0050154.sg002 (16 KB EPS).

Table S1. References for the Expression Data Used

Found at doi:10.1371/journal.pbio.0050154.st001 (28 KB DOC).

Text S1. Testing the False Negative Rate of the Dip Test

Found at doi:10.1371/journal.pbio.0050154.sd001 (20 KB DOC).

Text S2. No Difference in the Rate of Evolution of Singlish and Multi-Proteins when Controlling for Abundance

Found at doi:10.1371/journal.pbio.0050154.sd002 (25 KB DOC).

Dataset S1. Updated High-Confidence Interaction Dataset

Found at doi:10.1371/journal.pbio.0050154.sd003 (2 MB TXT)

References

1. Han JD, Bertin N, Hao T, Goldberg DS, Berriz GF, et al. (2004) Evidence for dynamically organized modularity in the yeast protein-protein interaction network. Nature 430: 88–93.
2. Batada NN, Reguly T, Breitkre tz A, Boucher L, Breitkreutz BJ, et al. (2006) uStratus not altocumulus: A new view of the yeast protein interaction network. PLoS Biol 4(10): e317 doi:10.1371/journal.pbio.0040317.
3. Bertin N, Simonis N, Dupuy D, Cusick ME, Han JDJ, et al. (2007) Confirmation of organized modularity in the yeast interactome. PLoS Biol 5(6): e153. doi:10.1371/journal.pbio.0050153.
4. Fraser HB (2005) Modularity and evolutionary constraint on proteins. Nat Genet 37: 351–352.
5. Gavin AC, Bosche M, Krause R, Grandi P, Marzioch M, et al. (2002) Functional organization of the yeast proteome by systematic analysis of protein complexes. Nature 415: 141–147.
6. Gavin AC, Aloy P, Grandi P, Krause R, Boesche M, et al. (2006) Proteome survey reveals modularity of the yeast cell machinery. Nature 440: 631–636.
7. Hartigan JA, Hartigan PM (1985) The dip test of unimodality. Ann Stat 13: 70–84.
8. Hartigan PM (1985) Computation of the dip statistic to test for unimodality. J Roy Stat Soc C, App Stat 34: 320–325.
9. Drummond DA, Bloom JD, Adami C, Wilke CO, Arnold FH (2005) Why highly expressed proteins evolve slowly. Proc Natl Acad Sci U S A 102: 14338–14343.
10. Hirsh AE, Fraser HB, Wall DP (2005) Adjusting for selection on synonymous sites in estimates of evolutionary distance. Mol Biol Evol 22: 174–177.
11. Newman JR, Ghaemmaghami S, Ihmels J, Breslow DK, Noble M, et al. (2006) Single-cell proteomic analysis of *S. cerevisiae* reveals the architecture of biological noise. Nature 441: 840–846.
12. R Development Core Team (2005) R: A language and environment for statistical computing. Vienna (Austria): R Foundation for Statistical Computing.

13. Kim PM, Lu LJ, Xia Y, Gerstein MB (2006) Relating three-dimensional structures to protein networks provides evolutionary insights. Science 314: 1938–1941.
14. Batada NN, Hurst LD, Tyers M (2006) Evolutionary and physiological importance of hub proteins. PLoS Comp Bio 2(7): e88 doi:10.1371/journal.pcbi.0020088.
15. Reguly T, Breitkreutz A, Boucher L, Breitkreutz B-J, Hon G, et al. (2006) Comprehensive curation and analysis of global interaction networks in Saccharomyces cerevisiae. J Biol 5: 11.
16. Ozier O, Amin N, Ideker T (2003) Global architecture of genetic interactions on the protein network. Nat Biotechnol 21: 490–491.

Citation: Batada NN, Reguly T, Breitkreutz A, Boucher L, Breitkreutz BJ, et al. (2007) Still stratus not altocumulus: Further evidence against the date/party hub distinction. PLoS Biol 5(6): e154. doi:10.1371/journal.pbio.0050154

Nizar N. Batada and Laurence D. Hurst are with the Department of Biology and Biochemistry, University of Bath, Bath, United Kingdom. Teresa Reguly, Ashton Breitkreutz, Lorrie Boucher, Bobby-Joe Breitkreutz, and Mike Tyers are with Samuel Lunenfeld Research Institute, Mount Sinai Hospital, Toronto, Canada. Lorrie Boucher and Mike Tyers are also with the Department of Medical Genetics and Microbiology, University of Toronto, Toronto, Canada.

* To whom correspondence should be addressed. E-mail: nizar.batada@gmail.com (NNB); l.d.hurst@bath.ac.uk (LDH); tyers@mshri.on.ca (MT)