Science
AAAS

High-Quality Binary Protein Interaction Map of the Yeast Interactome Network
Haiyuan Yu, *et al.*
Science **322**, 104 (2008);
DOI: 10.1126/science.1158684

Science (print ISSN 0036-8075; online ISSN 1095-9203) is published weekly, except the last week in December, by the American Association for the Advancement of Science, 1200 New York Avenue NW, Washington, DC 20005.

despite their socioeconomic importance. Until now, the selection of genomes for sequencing has been determined on the basis of genome simplicity and not agronomic relevance, with serious consequences for crop improvement and food security [for example, by neglecting wheat or choosing the diploid of cotton, *Gossypium raimondii*, to sequence first rather than focusing on the economically important tetraploid *G. hirsutum* (*24*)]. Our work may pave the way for a major change in how the next genomes for de novo sequencing are selected, thereby accelerating improvements in economically important crop species.

References and Notes

1. B. Wollenweber *et al.*, *Curr. Opin. Plant Biol.* **8**, 337 (2005).
2. T. Itoh *et al.*, *Genome Res.* **17**, 175 (2007).
3. R. L. Warren *et al.*, *Genome Res.* **16**, 768 (2006).
4. M. Chen *et al.*, *Plant Cell* **14**, 537 (2002).
5. P. E. Klein *et al.*, *Genome Res.* **10**, 789 (2000).
6. F. Wei *et al.*, *PLoS Genet.* **3**, e123 (2007).
7. M. C. Luo *et al.*, *Genomics* **82**, 378 (2003).
8. J. Dolezel *et al.*, *Chromosome Res.* **15**, 51 (2007).
9. J. Safar *et al.*, *Plant J.* **39**, 960 (2004).
10. T. R. Endo, B. S. Gill, *J. Hered.* **87**, 295 (1996).
11. Materials and methods are available as supporting material on *Science* Online.
12. B. S. Gill *et al.*, *Genome* **34**, 830 (1991).
13. E. Paux *et al.*, *Plant J.* **48**, 463 (2006).
14. J. D. Munkvold *et al.*, *Genetics* **168**, 639 (2004).
15. M. Akbari *et al.*, *Theor. Appl. Genet.* **113**, 1409 (2006).
16. M. Dilbirligi *et al.*, *Genetics* **166**, 461 (2004).
17. K. G. Hossain *et al.*, *Genetics* **168**, 415 (2004).
18. K. C. Cone *et al.*, *Plant Physiol.* **130**, 1598 (2002).
19. M. La Rota, M. E. Sorrells, *Funct. Integr. Genomics* **4**, 34 (2004).
20. R. Kota *et al.*, *Theor. Appl. Genet.* **112**, 492 (2006).
21. S. Liu *et al.*, *Funct. Integr. Genomics* **6**, 83 (2006).
22. M. D. McNeil *et al.*, *Mol. Breed.* **22**, 15 (2008).
23. J. Dolezel *et al.*, *Chromosome Res.* **12**, 77 (2004).
24. Z. J. Chen *et al.*, *Plant Physiol.* **145**, 1303 (2007).

25. We thank D. Boyer, K. Paux, and members of the GENTYANE platform for technical assistance; M. Ganal (Trait Genetics), P. Jack (RAGT), and F. Azanza (Syngenta) for providing SSR markers and genotyping data; the support of F. Legeai and D. Steinbach from Unité de Recherche en Génomique-Info in establishing the Gbrowse display of the integrated physical map; A. Kumar (Choudhury Charan Singh University Meerut), O. Riera-Lizarazu (Oregon State University), and Y. Gu (U.S. Department of Agriculture–Agricultural Research Service) for significant contributions to the RH mapping work; and F. Quetier (Agence Nationale de la Recherche) for support and discussions. Supported by the Agence Nationale de la Recherche (grant ANR-05-BLANC-0258-01) and the INRA.

Supporting Online Material

www.sciencemag.org/cgi/content/full/322/5898/101/DC1
Materials and Methods
Figs. S1 to S3
Tables S1 to S8
References

16 June 2008; accepted 10 September 2008
10.1126/science.1161847

High-Quality Binary Protein Interaction Map of the Yeast Interactome Network

Haiyuan Yu,[1,2]* Pascal Braun,[1,2]* Muhammed A. Yıldırım,[1,2,3]* Irma Lemmens,[4] Kavitha Venkatesan,[1,2] Julie Sahalie,[1,2] Tomoko Hirozane-Kishikawa,[1,2] Fana Gebreab,[1,2] Na Li,[1,2] Nicolas Simonis,[1,2] Tong Hao,[1,2] Jean-François Rual,[1,2] Amélie Dricot,[1,2] Alexei Vazquez,[5] Ryan R. Murray,[1,2] Christophe Simon,[1,2] Leah Tardivo,[1,2] Stanley Tam,[1,2] Nenad Svrzikapa,[1,2] Changyu Fan,[1,2] Anne-Sophie de Smet,[4] Adriana Motyl,[6] Michael E. Hudson,[6] Juyong Park,[1,7] Xiaofeng Xin,[8] Michael E. Cusick,[1,2] Troy Moore,[9] Charlie Boone,[8] Michael Snyder,[6] Frederick P. Roth,[1,10] Albert-László Barabási,[1,7] Jan Tavernier,[4] David E. Hill,[1,2] Marc Vidal[1,2]†

Current yeast interactome network maps contain several hundred molecular complexes with limited and somewhat controversial representation of direct binary interactions. We carried out a comparative quality assessment of current yeast interactome data sets, demonstrating that high-throughput yeast two-hybrid (Y2H) screening provides high-quality binary interaction information. Because a large fraction of the yeast binary interactome remains to be mapped, we developed an empirically controlled mapping framework to produce a "second-generation" high-quality, high-throughput Y2H data set covering ~20% of all yeast binary interactions. Both Y2H and affinity purification followed by mass spectrometry (AP/MS) data are of equally high quality but of a fundamentally different and complementary nature, resulting in networks with different topological and biological properties. Compared to co-complex interactome models, this binary map is enriched for transient signaling interactions and intercomplex connections with a highly significant clustering between essential proteins. Rather than correlating with essentiality, protein connectivity correlates with genetic pleiotropy.

A crucial step toward understanding cellular systems properties is mapping networks of physical DNA-, RNA-, and protein-protein interactions, the "interactome network," of an organism of interest as completely and accurately as possible. One approach consists in systematically testing all pairwise combinations of predicted proteins to derive the "binary" interactome. Early attempts at binary interactome mapping used high-throughput yeast two-hybrid (Y2H) screening, in which a protein interaction reconstitutes a transcription factor that activates expression of reporter genes. High-throughput Y2H maps have been generated for *Saccharomyces cerevisiae* (*1*–*3*), *Caenorhabditis elegans* (*4*–*6*), *Drosophila melanogaster* (*7*), and humans (*8*–*10*). An alternative approach consists in generating "co-complex" interactome maps, achievable by high-throughput coaffinity purification followed by mass spectrometry (AP/MS) to identify proteins bound to tagged baits, as done for *Escherichia coli* (*11*, *12*), *S. cerevisiae* (*13*–*16*), and humans (*17*).

To investigate fundamental questions of interactome network structure and function, it is necessary to understand how the size and quality of currently available maps, including thorough evaluation of differences between binary and co-complex maps, might have affected conclusions about global and local properties of interactome networks (*18*, *19*). Here, we address these issues using the yeast *S. cerevisiae* as a model system.

First, we compared the quality of existing high-throughput binary and co-complex data sets to information obtained from curating low-throughput experiments described in the literature (Fig. 1A). For binary interactions, we examined (i) the subset found by Uetz *et al.* in a proteome-scale all-by-all screen ("Uetz-screen"), excluding the pairs found in a focused, potentially biased experiment involving only 193 baits ("Uetz-array") (*2*); and (ii) the Ito *et al.* interactions found three times or more ("Ito-core"), independently from those found one or two times ("Ito-noncore"), a distinction recommended by the authors but seldom applied in the literature (*3*). For co-complex associations, we investigated two high-throughput AP/MS data sets referred to as "Gavin" (*15*) and "Krogan" (*16*). For literature-curated interactions, we considered only those curated from two or more publications ("LC-multiple") (*20*), which we judged of higher quality than those curated from a single publication.

To experimentally compare the quality of these data sets, we selected a representative sample of ~200 protein interaction pairs from each one and tested them by means of two independent interaction assays, Y2H and a yellow fluorescent protein complementation assay (PCA) (*21*) [Supporting Online Material (SOM) I]. In PCA, bait and prey proteins are fused to nonfluorescent fragments of yellow fluorescent protein that, when brought in close proximity by interacting proteins, reconstitute a fluorescent protein in mammalian cells. In con-

trast, reconstitution of a transcription factor in Y2H experiments takes place in the nucleus of yeast cells. In terms of assay designs, Y2H and PCA can be considered as orthogonal assays and can be used to validate each other's results.

No single assay is expected to detect 100% of genuine interactions, and the actual fraction of positives detected is inherently linked to the stringency at which the assay is implemented. To identify the optimal scoring condition of each assay, we selected a set of ~100 well-documented yeast protein-protein interaction pairs ["positive reference set" (PRS)] and a set of ~100 random pairs ["random reference set" (RRS)] (Fig. 1B; SOM II). Because RRS pairs were picked uniformly from the 14×10^6 possible pairings of proteins within our yeast ORFeome collection (*22*) (excluding those reported as interacting), these pairs are extremely unlikely to be interacting.

Sampled pairs from binary Uetz-screen and Ito-core data sets tested positive at levels as high as those of the positive-control PRS, demonstrating their high quality (Fig. 1C). A sample of literature-curated LC-multiple interactions tested slightly lower with Y2H, while being indistinguishable by PCA (Fig. 1C), demonstrating that high-throughput Y2H data sets can be comparable in quality to literature-curated information. In marked contrast, sampled pairs from Ito-noncore tested positive at levels similar to those of the negative-control RRS, confirming the low quality of this particular data set (Fig. 1C).

Sampled pairs from Gavin and Krogan high-throughput AP/MS data sets tested poorly in our two binary interaction assays (Fig. 1C), albeit at levels similar to those of Munich Information Center for Protein Sequences (MIPS) complexes, a widely used "gold standard" (*23*). This observation demonstrates that, at least for detecting binary interactions, Y2H performs better than AP/MS.

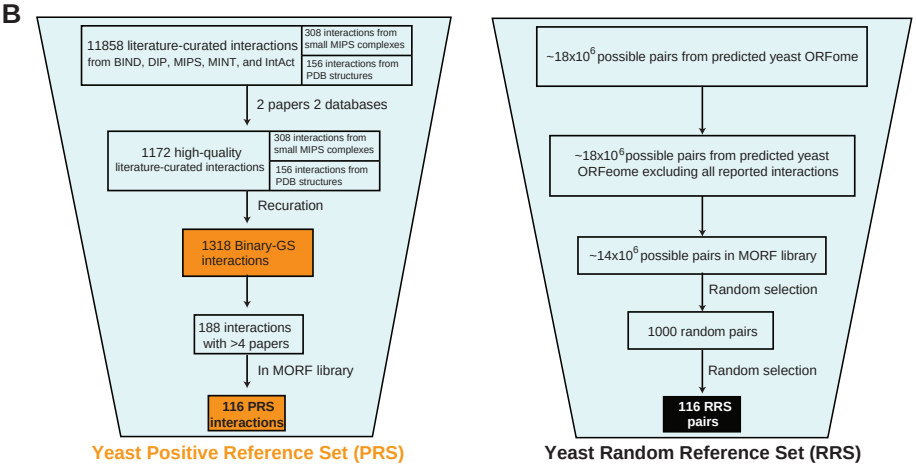
Our experimental data quality assessment shows that binary Uetz-screen, Ito-core, and LC-multiple data sets are of high quality, whereas Ito-noncore should not be used. AP/MS data sets, although of intrinsically good quality (*15*, *16*), should be used cautiously when binary interaction information is needed.

Our experimental results contrast markedly with computational analyses that suggested that high-throughput Y2H data sets contain more false-positives than literature-curated or high-throughput AP/MS data sets (*24*, *25*). In computational analyses, the quality of a data set is often determined by the fraction of interactions also present in a predefined gold standard set (*24*). Generally, MIPS complexes have been considered as gold standard, with all proteins constituting a given complex modeled as interacting with each other. Such modeling results in limited and biased sampling issues against binary interactions because not all

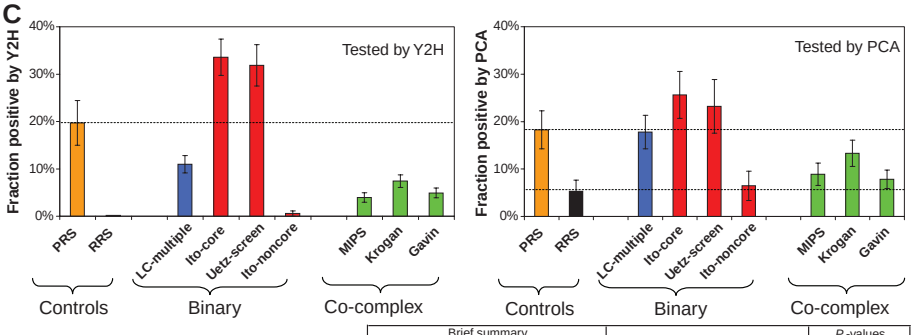
A

Reference	Uetz *et al.* (*2*)		Ito *et al.* (*3*)		Gavin *et al.* (*15*)	Krogan *et al.* (*16*)	Reguly *et al.* (*20*)
Dataset name	Uetz-screen	Uetz-array	Ito-core	Ito-noncore	Gavin	Krogan	LC-multiple
Number of reported interactions	682	281	843	3632	6942	7123	2925
	Binary				Co-complex		Literature

B

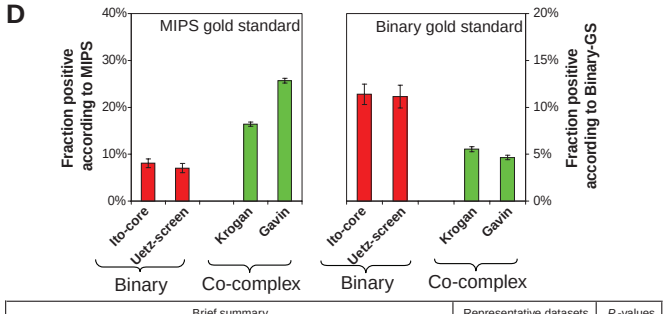


C



Brief summary (For detecting binary interactions)	Representative datasets	*P*-values Y2H	*P*-values PCA
1. Y2H is comparable to LC-multiple	Uetz-screen ≥ LC-multiple	$P<10^{-16}$	$P=0.1$
2. Ito-noncore is of extremely low quality	Ito-noncore ≈ RRS	$P=0.9$	$P=0.2$
3. Y2H performs better than AP/MS	Uetz-screen > Krogan	$P<10^{-30}$	$P<10^{-3}$
4. Y2H performs better than MIPS	Uetz-screen > MIPS	$P<10^{-41}$	$P<10^{-5}$

D



Brief summary	Representative datasets	*P*-values
1. AP/MS performs well at detecting co-complex associations according to MIPS	Uetz-screen < Krogan	$P<10^{-12}$
2. Y2H performs well at detecting binary interactions according to Binary-GS	Uetz-screen > Krogan	$P<10^{-5}$

Fig. 1. Evaluation of *S. cerevisiae* protein-protein interaction data sets. (**A**) Number of interactions reported in various large-scale *S. cerevisiae* protein-protein interaction data sets. (**B**) Schema of pipeline used to assemble binary positive and random reference sets. (**C**) Fraction of a random sample of interactions from each data set confirmed by Y2H and PCA. (**D**) Fraction of positive interactions in each data set calculated with MIPS and Binary-GS. Error bars indicate SE.

[1]Center for Cancer Systems Biology (CCSB), Dana-Farber Cancer Institute, Boston, MA 02115, USA. [2]Department of Cancer Biology, Dana-Farber Cancer Institute, and Department of Genetics, Harvard Medical School, Boston, MA 02115, USA. [3]School of Engineering and Applied Sciences, Harvard University, Cambridge, MA 02138, USA. [4]Department of Medical Protein Research, VIB, and Department of Biochemistry, Faculty of Medicine and Health Sciences, Ghent University, 9000 Ghent, Belgium. [5]The Simons Center for Systems Biology, Institute for Advanced Studies, Princeton, NJ 08540, USA. [6]Department of Molecular, Cellular, and Developmental Biology, Yale University, New Haven, CT 06620, USA. [7]Center for Complex Network Research and Departments of Physics, Biology, and Computer Science, Northeastern University, Boston, MA 02115, USA. [8]Banting and Best Department of Medical Research and Department of Medical Genetics and Microbiology, Terrence Donnelly Centre for Cellular and Biomolecular Research, University of Toronto, Toronto, ON M5S 3E1, Canada. [9]Open Biosystems, Huntsville, AL 35806, USA. [10]Department of Biological Chemistry and Molecular Pharmacology, Harvard Medical School, Boston, MA 02115, USA.

*These authors contributed equally to this work.
†To whom correspondence should be addressed. E-mail: marc_vidal@dfci.harvard.edu

proteins in a complex contact each other directly (fig. S1), and not all direct physical interactions occur within complexes (fig. S2 and SOM III). Hence, although MIPS complexes are appropriate for benchmarking co-complex membership data sets, they are not appropriate for binary interaction data sets. This distinction is corroborated by the poor experimental confirmation rate of pairs from MIPS complexes with binary assays (Fig. 1C).

To computationally reexamine the quality of existing yeast interactome data sets, we assembled a binary gold standard set ("Binary-GS") of 1318 high-confidence physical binary interactions (Fig. 1B and SOM III). Binary-GS includes direct physical interactions within well-established complexes, as well as conditional interactions (e.g., dependent on posttranslational modifications), and thus represents well-documented direct physical interactions in the yeast interactome (*26*). When measured against Binary-GS, the quality of high-throughput Y2H data sets (with the exception of Ito-noncore) was substantially better (SOM IV and V) than that of high-throughput AP/MS data sets (Fig. 1D). Our results demonstrate the distinct nature of binary and co-complex data. Generally, Y2H data sets contain high-quality direct binary interactions, whereas AP/MS co-complex data sets are composed of direct interactions mixed with preponderant indirect associations (SOM VI).

The proteome-wide binary data sets, Uetz-screen and Ito-core, contain 682 and 843 interactions, respectively (*2*, *3*). The overlap between these two data sets appears low (*3*, *24*): 19% of Uetz-screen and 15% of Ito-core interactions were detected in the other data set. Given our demonstration of high quality for these data sets (Fig. 1, C and D), we conclude that the small overlap stems primarily from low sensitivity (i.e., many false-negatives) rather than from low specificity (i.e., many false-positives, as previously suggested).

Several factors might affect sensitivity. First, the space of pairwise protein combinations actually tested in each data set might have been considerably different. We refer to the fraction of all possible pairs tested in a given screen as the "completeness." For example, missing 10% of open reading frames (ORFs) in each mapping project could reduce the common tested space down to 66% [(0.9 × 0.9) × (0.9 × 0.9)] of all possible pairwise combinations. Second, different protein interaction assays or even different versions of the same assay detect different subsets of pairs out of all possible interactions, which explains partly the limited overlap between data sets obtained with different Y2H versions. For any assay, the "assay sensitivity" is estimated as the fraction of PRS interactions detected, which for our Y2H assay was determined empirically to be ~20% (Fig. 1C). Finally, when screening tens if not hundreds of millions of protein pairs in any tested space, that search space might need to be sampled multiple times to report all or nearly all interactions detectable by the assay used. The fraction of all theoretically detectable interactions by a particular assay found in a given experiment is its "sampling sensitivity." These three parameters fully account for the seemingly small overlap between Ito-core and Uetz-screen (SOM VII), demonstrating that a large fraction of the *S. cerevisiae* binary interactome remains to be mapped. Therefore, we carried out a new proteome-scale yeast high-throughput Y2H screen (fig. S3).

We used 5796 Gateway-cloned ORFs available in the yeast movable ORF (MORF) collection (*22*). After subcloning these ORFs into Y2H vectors and removing autoactivators (*27*, *28*), our search space became 3917 DB-Xs against 5246 AD-Ys, representing a completeness of 77% (Fig. 2A and SOM VI), comparable to that of recent AP/MS data sets (*15*) (~78%; SOM VI).

To address sampling sensitivity, we determined what fraction of all detectable interactions is found in each pass after eight trials in a search space of 658 DB-X and 1249 AD-Y ORFs. A single trial identified about 60% of all possible interactions that can be detected with our high-throughput Y2H, whereas three to five repeats were required to obtain 80 to 90% (Fig. 2B and SOM VI). Consequently, we screened the whole search space three times independently to yield an estimated sampling sensitivity of 85% (Fig. 2B). In total, ~88,000 colonies were selected, of which 21,432 scored positive upon more detailed phenotyping (SOM I). After identifying all putative interaction pairs by sequencing, phenotypically retesting them with fresh cultures from archival stocks, and eliminating de novo autoactivators (*28*), we obtained a final data set, "CCSB-YI1," of 1809 interactions among 1278 proteins, which can be downloaded from our Web site (http://interactome.dfci.harvard.edu/S_cerevisiae).

To validate the overall quality of CCSB-YI1, we tested 94 randomly chosen interactions by PCA and mammalian protein-protein interaction trap (MAPPIT; SOM I) (*21*, *29*). MAPPIT takes place at the mammalian cell membrane and measures interactions via activation of signal transducer and activator of transcription 3 (STAT3)–dependent reporter expression. Using both PCA and MAPPIT, we found that the confirmation rate of CCSB-YI1 was similar to those of Ito-core and Uetz-screen (Fig. 1C). The precision [i.e., fraction of true positives in the data set (*30*)] of CCSB-YI1 is estimated at 94 to 100% (Fig. 2C, fig. S4, and SOM VI). Additionally, the performance of our high-throughput Y2H approach was confirmed via a larger RRS of 1000 random pairs (*30*) (Fig. 1B), none of which tested positive (SOM II).

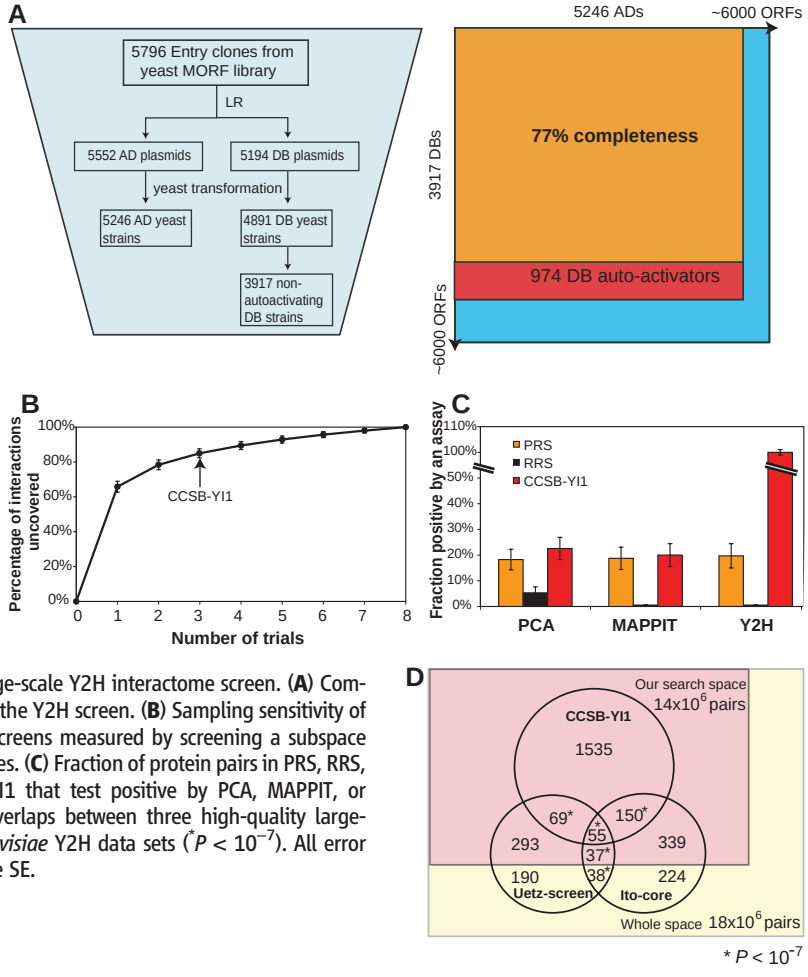


Fig. 2. Large-scale Y2H interactome screen. (**A**) Completeness of the Y2H screen. (**B**) Sampling sensitivity of CCSB Y2H screens measured by screening a subspace multiple times. (**C**) Fraction of protein pairs in PRS, RRS, and CCSB-YI1 that test positive by PCA, MAPPIT, or Y2H. (**D**) Overlaps between three high-quality large-scale *S. cerevisiae* Y2H data sets ($^*P < 10^{-7}$). All error bars indicate SE.

The overlaps of Uetz-screen (27%) and Ito-core (35%) with CCSB-YI1 (Fig. 2D) can be explained by the completeness, assay sensitivity, and sampling sensitivity of the three experiments (SOM VII) and agree well with the results of the pairwise confirmation of those two data sets (Fig. 1C). Similar principles apply to other large-scale experiments such as AP/MS, likely accounting for the low overlap between Krogan and Gavin (~25%; fig. S5B).

Factoring in completeness, precision, and assay and sampling sensitivity, we estimated that the yeast binary interactome consists of ~18,000 ± 4500 interactions (SOM VI), experimentally validating previous computational estimates of 17,000

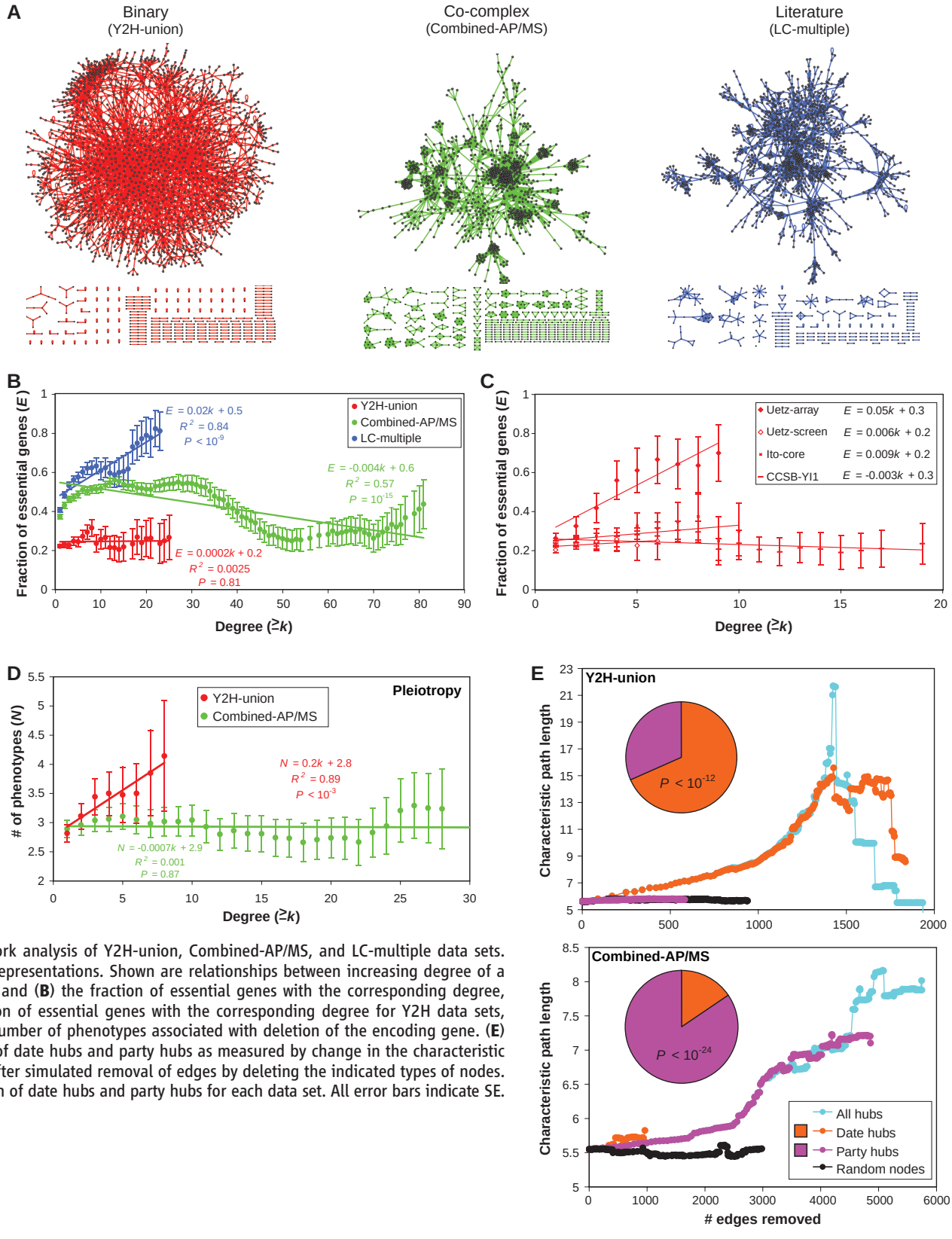


Fig. 3. Network analysis of Y2H-union, Combined-AP/MS, and LC-multiple data sets. (**A**) Network representations. Shown are relationships between increasing degree of a gene product and (**B**) the fraction of essential genes with the corresponding degree, (**C**) the fraction of essential genes with the corresponding degree for Y2H data sets, and (**D**) the number of phenotypes associated with deletion of the encoding gene. (**E**) Contribution of date hubs and party hubs as measured by change in the characteristic path length after simulated removal of edges by deleting the indicated types of nodes. (Inset) Fraction of date hubs and party hubs for each data set. All error bars indicate SE.

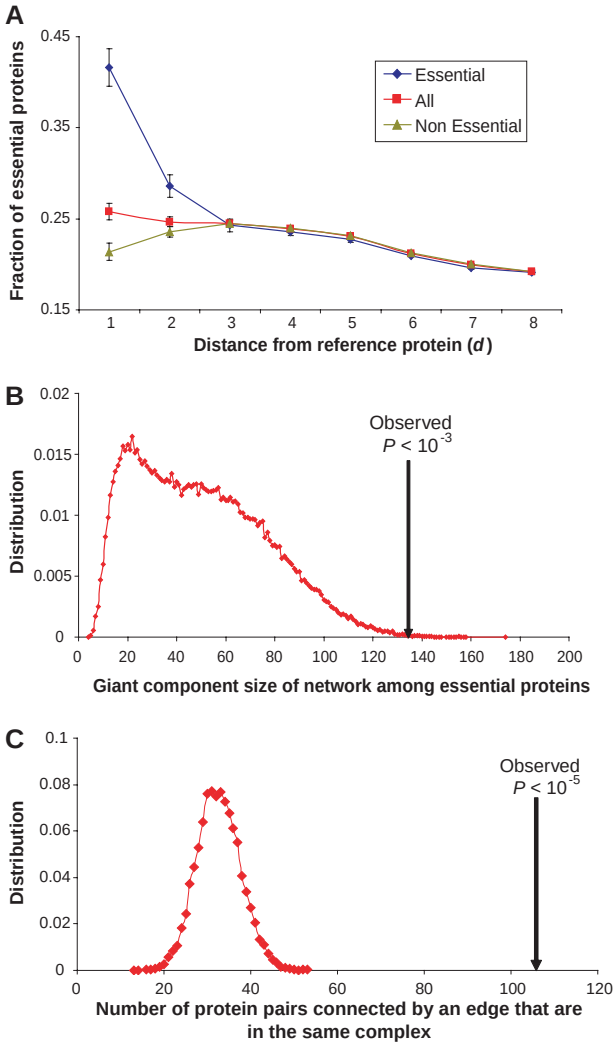
to 25,000 interactions (*31*, *32*). To obtain a more comprehensive map of the binary yeast interactome, we combined the three available high-quality proteome-scale Y2H data sets (SOM VII). The union of Uetz-screen, Ito-core, and CCSB-YI1, "Y2H-union," contains 2930 binary interactions among 2018 proteins, which, according to our empirical estimate of the interactome size, represents ~20% of the whole yeast binary interactome (Fig. 3A).

We reexamined global topological features of this new yeast interactome network, facing lower risk of overinterpreting properties due to limited sampling and various biases in the data (*18*). To contrast topological properties of the binary Y2H-union network with that of the co-complex network, we used an integrated AP/MS data set (*33*), which was generated by combining raw high-throughput AP/MS data (*15*, *16*). This "Combined-AP/MS" data set, composed of 9070 co-complex membership associations between 1622 proteins, attempts to model binary interactions from co-complex data (Fig. 3A).

As found previously for other macromolecular networks, the connectivity or "degree" distribution of all three data sets is best approximated by a power-law (*34*) (fig. S6 and SOM VIII). Highly connected proteins, or "hubs," are reportedly more likely encoded by essential genes than less-connected proteins (*35*). Surprisingly, Y2H-union lacked any correlation between degree and essentiality (Fig. 3B). This discrepancy might stem from biases in the data sets available at the time of the original observation: interactions reported in Uetz *et al*. (Uetz-array and Uetz-screen) and literature-curated interactions. Although Uetz-array is of high quality (fig. S7), its experimental design could negatively influence network analyses. Most hub proteins in Uetz-array were found as baits (fig. S8), and the percentage of essential proteins in the 193 bait proteins is twice as high (34.7%) as that of all protein-encoding ORFs in the yeast genome (18.4%), explaining the high correlation between degree and essentiality (Fig. 3C). Likewise, literature-curated interactions seem prone to sociological and other inspection biases (SOM VII). Thus, we refrain hereinafter from using LC-multiple in our further topological and biological analyses. No significant correlation between degree of connectedness and essentiality was observed in any of the three proteome-wide high-throughput binary data sets currently available (i.e., Ito-core, Uetz-screen, and CCSB-YI1; Fig. 3C), as well as in new versions of our *C. elegans* and human interactome maps (fig. S9 and SOM IX).

Hub proteins instead relate to pleiotropy, the number of phenotypes observed as a consequence of gene knockout (SOM I). There was a significant correlation in Y2H-union between connectivity and the number of phenotypes observed in global phenotypic profiling analyses of yeast genes (*36*) (Fig. 3D). Thus, the number of binary physical interactions mediated by a protein seems to better correlate with the number of cellular processes in which it participates than with its essentiality. The correlation between degree and number of phenotypes is not observed in Combined-AP/MS, likely because co-complex associations reflect the size of protein complexes more than the number of processes they might be involved in.

Fig. 4. Clustering of essential proteins. (**A**) Average fraction of essential proteins among proteins whose distance is equal to *d* from a protein selected from essential, nonessential, and all proteins. (**B**) Giant component size of network formed by essential proteins (arrow) compared to 100,000 random networks of same topological properties. (**C**) The number of interacting essential proteins that are also found in the same complex compared to 10,000 random selections of proteins of the same number as essential proteins (SOM IX).



We confirmed the concept of modularity in the yeast interactome network, whereby date hubs that dynamically interact with their partners appear particularly central to global connectivity, whereas static party hubs appear to function locally in specific biological modules (*37*). The proportion of date and party hubs is substantially different between Y2H-union and Combined-AP/MS (Fig. 3E). There are significantly more date hubs in the binary network, whereas party hubs are prevalent in the co-complex network. In the binary network, date hubs are crucial to the topological integrity of the network, whereas party hubs have minimal effects. However, in the co-complex network, date and party hubs affect the topological integrity of the network equally, likely because most hubs in Combined-AP/MS reside in large stable complexes, whereas hubs in Y2H-union preferentially connect diverse cellular processes.

Surprisingly, essential proteins strongly tended to interact with each other (Fig. 4A and SOM IX). By concentrating on the subnetwork formed by interactions mediated by and among essential proteins (fig. S10), we found a giant component whose size is much larger than expected by chance (Fig. 4B). To better understand the clustering of essential proteins, we examined the interacting essential protein pairs that are also reported to be in the same complex; we found 106 interacting essential protein pairs, a number greater than expected by chance (Fig. 4C and SOM IX).

We investigated the overall relationships between Y2H-union and Gene Ontology (GO) attributes (*38*), phenotypic and expression profiling similarities (*39*), and transcriptional regulatory networks (*40*). Both Y2H-union and Combined-AP/MS show significant enrichment (all $P < 10^{-10}$) for functionally similar pairs in all three GO branches (Fig. 5A) (*41*). There is also significant enrichment of positive correlations of phenotypic profiles (*36*) between interacting pairs in both data sets (Fig. 5B and fig. S11). Such interactions, when supported by strong phenotypic information, constitute likely possibilities of functional relationships. Lastly, both data sets are significant-

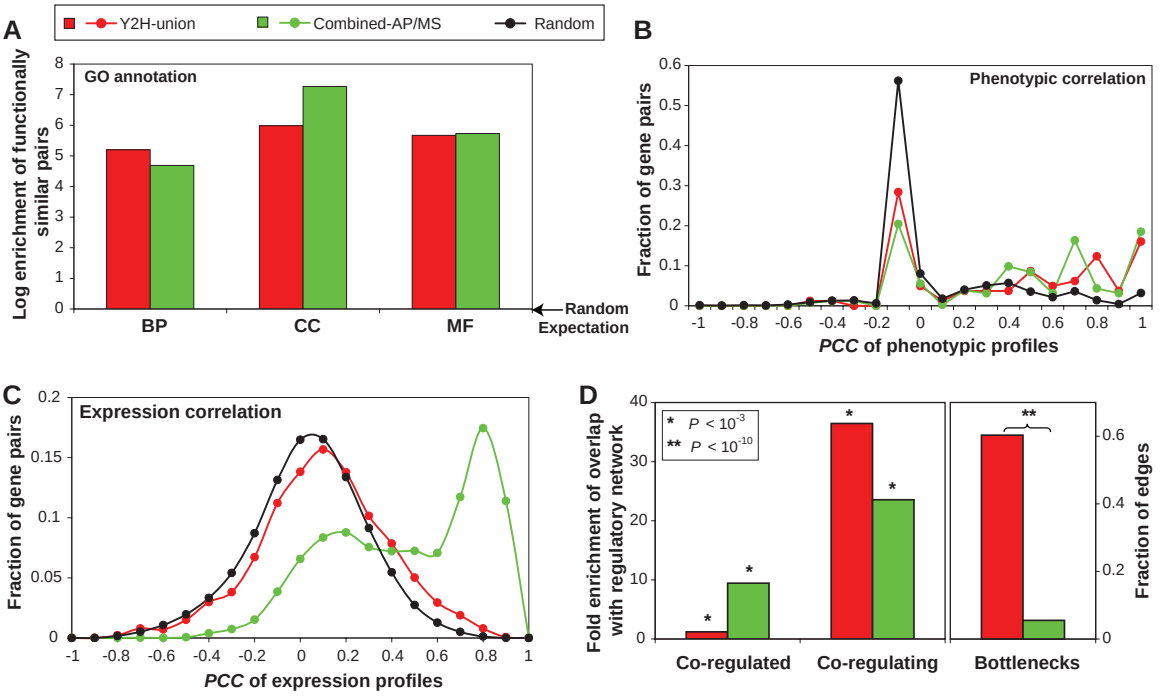


Fig. 5. Biological features of yeast interactome data sets. (**A**) Enrichment of interacting protein pairs (relative to random) that share GO annotations in the biological process, cellular component, and molecular function branches of GO ontology. (**B**) Pearson correlation coefficient (PCC) of phenotypic profiles between interacting pairs in different data sets. (**C**) Coexpression correlation between interacting pairs. (**D**) (Left) Enrichment of interacting proteins as targets of a common TF (co-regulated), and enrichment of interacting TFs in a common MIM (co-regulating) (*$P < 10^{-3}$). (Right) Fraction of bottlenecks from each data set in the combined network (SOM XI). Top 10% of edges with the highest betweenness are defined as "bottlenecks" (*45*).

ly enriched with pairs coexpressed across many conditions (fig. S12), although Combined-AP/MS shows higher enrichment (Fig. 5C), agreeing well with the different nature of the two assays: AP/MS aims at detecting stable complexes, whereas Y2H tends to detect more transient and condition-specific protein interactions. This observation is further supported by enrichment of kinase-substrate pairs in Y2H-union (SOM X and fig. S13).

To explore the mechanisms underlying coexpression of interacting protein pairs, we combined transcriptional regulatory networks with interactome network information (*40*). Interacting proteins in both networks tended to be co-regulated by common transcription factors (TFs; Fig. 5D). Similar to what we observed in the coexpression correlation analysis (Fig. 5C), the enrichment of interacting pairs in Combined-AP/MS was significantly higher than that of Y2H-union. Notably, we observed a significant enrichment of protein-protein interactions between TFs involved in a common "multi-input motif" (*42*, *43*) (MIM, where multiple TFs co-regulate a given set of genes; Fig. 5D and SOM X). The fraction of co-regulating TF pairs is much higher in the binary interactome than in the co-complex network, suggesting that various TFs function together to form transient complexes to differentially regulate transcriptional targets (*44*).

These observations suggest that our binary interactome data set is enriched in transient or condition-specific interactions linking different subcellular processes and molecular machines. To further explore this possibility, we calculated "edge-betweenness" for each interaction in a merged network of all available interactions (SOM XI), measuring the number of shortest paths between all protein pairs that traverse a given edge. The higher edge-betweenness of interactions from Y2H-union shows the tendency of Y2H to detect key interactions outside of complexes significantly more often than AP/MS (Fig. 5D). Several examples of such complex-to-complex connectivity are evident in a complete map of MIPS complexes connected by Y2H interactions (fig. S14).

Overall, we infer that Y2H interrogates a different subspace within the whole interactome than does AP/MS, and Y2H interactions represent key connections between different complexes and pathways. Y2H and AP/MS provide orthogonal information about the interactome and are both vital to obtaining a complete picture of cellular protein-protein interaction networks.

References and Notes

1. M. Fromont-Racine, J. C. Rain, P. Legrain, *Nat. Genet.* **16**, 277 (1997).
2. P. Uetz *et al.*, *Nature* **403**, 623 (2000).
3. T. Ito *et al.*, *Proc. Natl. Acad. Sci. U.S.A.* **98**, 4569 (2001).
4. A. J. Walhout *et al.*, *Science* **287**, 116 (2000).
5. J. Reboul *et al.*, *Nat. Genet.* **34**, 35 (2003).
6. S. Li *et al.*, *Science* **303**, 540 (2004).
7. L. Giot *et al.*, *Science* **302**, 1727 (2003).
8. F. Colland *et al.*, *Genome Res.* **14**, 1324 (2004).
9. J. F. Rual *et al.*, *Nature* **437**, 1173 (2005).
10. U. Stelzl *et al.*, *Cell* **122**, 957 (2005).
11. G. Butland *et al.*, *Nature* **433**, 531 (2005).
12. M. Arifuzzaman *et al.*, *Genome Res.* **16**, 686 (2006).
13. A. C. Gavin *et al.*, *Nature* **415**, 141 (2002).
14. Y. Ho *et al.*, *Nature* **415**, 180 (2002).
15. A. C. Gavin *et al.*, *Nature* **440**, 631 (2006).
16. N. J. Krogan *et al.*, *Nature* **440**, 637 (2006).
17. R. M. Ewing *et al.*, *Mol. Syst. Biol.* **3**, 89 (2007).
18. J. D. Han, D. Dupuy, N. Bertin, M. E. Cusick, M. Vidal, *Nat. Biotechnol.* **23**, 839 (2005).
19. D. Scholtens, M. Vidal, R. Gentleman, *Bioinformatics* **21**, 3548 (2005).
20. T. Reguly *et al.*, *J. Biol.* **5**, 11 (2006).
21. I. Remy, S. W. Michnick, *Methods Mol. Biol.* **261**, 411 (2004).
22. D. M. Gelperin *et al.*, *Genes Dev.* **19**, 2816 (2005).
23. H. W. Mewes *et al.*, *Nucleic Acids Res.* **34**, D169 (2006).
24. C. von Mering *et al.*, *Nature* **417**, 399 (2002).
25. J. S. Bader, A. Chaudhuri, J. M. Rothberg, J. Chant, *Nat. Biotechnol.* **22**, 78 (2004).
26. H. Yu *et al.*, *Genome Res.* **14**, 1107 (2004).
27. P. O. Vidalain, M. Boxem, H. Ge, S. Li, M. Vidal, *Methods* **32**, 363 (2004).
28. A. J. Walhout, M. Vidal, *Genome Res.* **9**, 1128 (1999).
29. S. Eyckerman *et al.*, *Nat. Cell Biol.* **3**, 1114 (2001).
30. R. Jansen, M. Gerstein, *Curr. Opin. Microbiol.* **7**, 535 (2004).
31. A. Grigoriev, *Nucleic Acids Res.* **31**, 4157 (2003).
32. R. Jansen *et al.*, *Science* **302**, 449 (2003).
33. S. R. Collins *et al.*, *Mol. Cell. Proteomics* **6**, 439 (2007).
34. A. L. Barabási, R. Albert, *Science* **286**, 509 (1999).
35. H. Jeong, S. P. Mason, A. L. Barabasi, Z. N. Oltvai, *Nature* **411**, 41 (2001).
36. A. M. Dudley, D. M. Janse, A. Tanay, R. Shamir, G. M. Church, *Mol. Syst. Biol.* **1**, 0001 (2005).
37. J. D. Han *et al.*, *Nature* **430**, 88 (2004).
38. Gene Ontology Consortium, *Nucleic Acids Res.* **36**, D440 (2008).
39. M. Vidal, *Cell* **104**, 333 (2001).
40. H. Yu, M. Gerstein, *Proc. Natl. Acad. Sci. U.S.A.* **103**, 14724 (2006).
41. H. Yu, R. Jansen, G. Stolovitzky, M. Gerstein, *Bioinformatics* **23**, 2163 (2007).
42. T. I. Lee *et al.*, *Science* **298**, 799 (2002).
43. R. Milo *et al.*, *Science* **298**, 824 (2002).
44. N. M. Luscombe *et al.*, *Nature* **431**, 308 (2004).
45. H. Yu, P. M. Kim, E. Sprecher, V. Trifonov, M. Gerstein, *PLOS Comput. Biol.* **3**, e59 (2007).

46. Supported by funds from the W. M. Keck Foundation (M.V. and F.P.R.); by Institute Sponsored Research funds from the Dana-Farber Cancer Institute Strategic Initiative (M.V. and CCSB); by NIH grant R01-HG001715 (M.V. and F.P.R.); by NIH grants U01-A1070499-01 and U56-CA113004 (A.-L.B.); by grant GOA12051401 from the University of Ghent and grant FWO-V G.0031.06 from the Fund for Scientific Research Flanders (J.T.); by a grant from the National Cancer Institute of Canada

(C.B.); and by NIH grant HG003224 (F.P.R.). I.L. is a postdoctoral fellow with the Fonds Wetenschappelijk Onderzoek–Vlanderen. M.V. is a "Chercheur Qualifié Honoraire" from the Fonds de la Recherche Scientifique (FRS-FNRS, French Community of Belgium). We thank members of our laboratories for helpful discussions and Agencourt Biosciences for sequencing assistance. All data sets can be downloaded from our Web site (http://interactome.dfci.harvard.edu/S_cerevisiae).

Supporting Online Material

www.sciencemag.org/cgi/content/full/1158684/DC1
SOM Text
Figs. S1 to S35
Tables S1 to S5
References

4 April 2008; accepted 1 August 2008
Published online 21 August 2008;
10.1126/science.1158684
Include this information when citing this paper.

Ceramide Biogenesis Is Required for Radiation-Induced Apoptosis in the Germ Line of *C. elegans*

Xinzhu Deng,[1] Xianglei Yin,[1] Richard Allan,[1] Diane D. Lu,[1] Carine W. Maurer,[2] Adriana Haimovitz-Friedman,[3] Zvi Fuks,[3] Shai Shaham,[2] Richard Kolesnick[1]*

Ceramide engagement in apoptotic pathways has been a topic of controversy. To address this controversy, we tested *loss-of-function (lf)* mutants of conserved genes of sphingolipid metabolism in *Caenorhabditis elegans*. Although somatic (developmental) apoptosis was unaffected, ionizing radiation–induced apoptosis of germ cells was obliterated upon inactivation of ceramide synthase and restored upon microinjection of long-chain natural ceramide. Radiation-induced increase in the concentration of ceramide localized to mitochondria and was required for BH3-domain protein EGL-1–mediated displacement of CED-4 (an APAF-1–like protein) from the CED-9 (a Bcl-2 family member)/CED-4 complex, an obligate step in activation of the CED-3 caspase. These studies define CEP-1 (the worm homolog of the tumor suppressor p53)–mediated accumulation of EGL-1 and ceramide synthase–mediated generation of ceramide through parallel pathways that integrate at mitochondrial membranes to regulate stress-induced apoptosis.

Although studies that use genetic deficiency in ceramide production support it as essential for apoptosis in diverse models (*1*), many have questioned whether ceramide functions as a bona fide transducer of apoptotic signals (*2*). One reason for skepticism is that, despite delineation of a number of ceramide-activated proteins, no single protein has been identified as mediator of ceramide-induced apoptosis. Recent studies have suggested an alternate mode of ceramide action, based on its capacity to self-associate and locally rearrange membrane bilayers into ceramide-rich macrodomains (1 to 5 μm in diameter), which are sites of protein concentration and oligomerization (*3*). Ceramide may thus mediate apoptosis through its ability to reconfigure membranes, coordinating protein complexation at critical junctures of signaling cascades.

To establish the role of ceramide definitively, we used a model of radiation-induced apoptosis in *Caenorhabditis elegans* germ cells (*4*). Germline stem cells, located at the distal gonad tip, divide incessantly throughout adult life, with daughter cells arresting in meiotic prophase. Upon exiting prophase, germ cells become sensitive to radiation-induced apoptosis, detected morphologically just proximal to the bend of the gonadal arm (*5*). This apoptotic pathway is antagonized by the ABL-1 tyrosine kinase, requiring sequentially the cell cycle checkpoint genes *rad-5*, *hus-1*, and *mrt-2*; the *C. elegans* p53 homolog *cep-1*; and the genes making up the conserved apoptotic machinery, the caspase *ced-3*, the apoptotic protease activating factor 1–like protein *ced-4*, the Bcl-2 protein *ced-9*, and the BH3-domain protein *egl-1*. This pathway differs from apoptotic somatic cell death, which is not subject to upstream checkpoint regulation via the CEP-1 pathway (*5*, *6*).

We identified conserved genes that regulate *C. elegans* sphingolipid intermediary metabolism and tested deletion alleles (Table 1 and table S1). Screening for mutants resistant to radiation-induced germ cell apoptosis revealed apoptosis suppression in only deletion mutants of *hyl-1* and *lagr-1*, two of the three ceramide synthase (CS) genes (Fig. 1A). CS gene products regulate de novo ceramide biosynthesis, acylating sphinganine to form dihydroceramide that is subsequently converted to ceramide by a desaturase (*7*). CSs contain six to seven putative transmembrane domains and a Lag1p motif [which confers enzyme activity (*8*)], regions conserved in the *C. elegans* orthologs. The deleted CS sequences in *hyl-1(ok976)* and *lagr-1(gk327)* result in frameshifts that disrupt the Lag1p motifs (fig. S1A). We detected a ~1.6-kb *hyl-1* transcript in wild-type (WT) worms and a smaller ~1.35-kb transcript in *hyl-1(ok976)*, whereas we observed a ~1.4-kb *lagr-1* transcript in WT worms and a ~1.25-kb transcript in *lagr-1(gk327)* (fig. S1B). In contrast, a deletion mutant of the third *C. elegans* CS (*9*, *10*), *hyl-2(ok1766)*, lacking a 1626-base pair fragment of the *hyl-2* gene locus that eliminates exons 2 to 5 corresponding to 74% of the coding sequence, displayed no defect in germ cell death (fig. S1C).

In N2 WT strain young adults, apoptotic germ cells gradually increased in abundance with age from a baseline of 0.7 ± 0.1 to 1.8 ± 0.2 corpses per distal gonad arm over 48 hours. Exposure to a 120-gray (Gy) ionizing radiation dose increased germ cell apoptosis to 5.2 ± 0.3 cells 36 to 48 hours after treatment. In contrast, in *hyl-1 (ok976)* and *lagr-1(gk327)* animals, age-dependent and radiation-induced germ cell apoptosis were nearly abolished (Fig. 1A). Similar effects were observed in the *lagr-1(gk327);hyl-1(ok976)* double mutant (Fig. 1B). The rate of germ cell corpse removal was unaffected in CS mutants, excluding the possibility that defective corpse engulfment elevated corpse numbers (table S2). In contrast, *loss-of-function (lf)* mutations of *hyl-1* or *lagr-1* did not affect developmental somatic cell death, nor did the *lf hyl-2(ok1766)* mutation (table S3). These studies indicate a requirement for two *C. elegans* CS genes for radiation-induced germline apoptosis.

To confirm ceramide as critical for germline apoptosis, we injected C_{16}-ceramide into gonads of young adult WT worms. C_{16}-ceramide is the predominant ceramide species in apoptosis induction by diverse stresses in multiple organisms (*11*) and in low abundance in *C. elegans* (*12*, *13*). C_{16}-ceramide microinjection resulted in time- and dose-dependent increases in germ cell apoptosis (Fig. 1C), with a median effective dose of ~0.05 μM gonadal ceramide. Peak effect occurred at ~0.1 μM gonadal ceramide at 36 hours (6.6 ± 0.8 versus 1.5 ± 0.4 cell corpses per distal gonad arm, $P < 0.0001$), qualitatively and quantitatively mimicking the 120-Gy effect in WT worms. In contrast, C_{16}-dihydroceramide, which differs from C_{16}-ceramide in a trans double bond at sphingoid base position four to five, was without effect (0.71 ± 0.28 cell corpses per distal gonad arm at ~1 μM), indicating specificity for ceramide in apoptosis induction. Furthermore, C_{16}-ceramide microinjection into *lagr-1(gk327);hyl-1(ok976)* animals (~1 μM gonadal ceramide) resulted in a 5.7-fold increase in germ cell apoptosis (from 0.60 ± 0.17 to 3.43 ± 0.88, $P < 0.0001$) (Fig. 1D). Note that the baseline level of apoptosis in *lagr-1(gk327);hyl-1(ok976)* was less than one-half that in WT worms. Moreover, ~0.005 μM gonadal ceramide, a concentration without impact on germ cell apoptosis, completely restored radiation (120 Gy)–induced apoptosis, an effect inhibitable in a *lf ced-3* background (Fig. 1E). C_{16}-ceramide's ability to bypass the genetic defect and restore the radiation-response pheno-

[1]Laboratory of Signal Transduction, Memorial Sloan-Kettering Cancer Center (MSKCC), New York, NY 10021, USA. [2]Laboratory of Developmental Genetics, Rockefeller University, New York, NY, 10021, USA. [3]Department of Radiation Oncology, Memorial Sloan-Kettering Cancer Center, New York, NY 10021, USA.

*To whom correspondence should be addressed. E-mail: r-kolesnick@ski.mskcc.org