A Map of the Interactome Network of the Metazoan C. elegans

Siming Li,1* Christopher M. Armstrong,1* Nicolas Bertin,1* Hui Ge,1* Stuart Milstein,1* Mike Bo xen,1* Pierre-Olivier Vidalain,1* Jing-Dong J. Han,1*
Alban Chesneau,1,2† Tong Hao,1† Debra S. Goldberg,3† Ning Li,1† Monica Martinez,1 Jean-François Rual,1,4† Philippe Lamesch,1,4† Lai Xu,5† Munee sh Tewari,1 Sharlyl L. Wong,1† Lan V. Zhang,3† Gabriel F. Berriz,5† Laurent Jacotot,1† Philippe Veglio,6† Jérôme Reboul,1† Tomoko Hirozane-Kishikawa,6,7 Qianru Li,1† Harrison W. Gabel,1† Ahmed Elewa,1† Bridget Baumgartner,5 Debra J. Rose,6 Haiyuan Yu,7 Stephanie Bosak,8 Reynaldo Sequerra,8 Andrew Fraser,9 Susan E. Margo,10 William M. Saxton,6 Susan Strome,6 Sander van den Heuvel,11 Fabio Piano,12 Jean Vandenhau te,4 Claude Sar det,2 Mark Gerstein,7 Lynn Doucette-Stamm,8 Kristin C. Gunsuals,12 J. Wade Harper,6,7 Michael E. Cuskic,1 Frederick P. Roth,9 David E. Hill,11 Marc Vidal4†

To initiate studies on how protein-protein interaction (or ‘interactome’) networks relate to multicellular functions, we have mapped a large fraction of the Caenorhabditis elegans interactome network. Starting with a subset of metazoan-specific proteins, more than 4000 interactions were identified from high-throughput, yeast two-hybrid (HT=Y2H) screens. Independent coefficiency purification assays experimentally validated the overall quality of this Y2H data set. Together with already described Y2H interactions and interologs predicted in silico, the current version of the Worm Interactome (WIS) map contains ~5500 interactions. Topological and biological features of this interactome network, as well as its integration with phenome and transcriptome data sets, lead to numerous biological hypotheses.

To further understand biological processes, it is important to consider protein functions in the context of complex molecular networks. The study of such networks requires the availability of proteome-wide protein-protein interaction, or ‘interactome,’ maps. The yeast Saccharomyces cerevisiae has been used to develop a eukaryotic unicellular interactome map (1–6). Caenorhabditis elegans is an ideal model for studying how protein networks relate to multicellularity. Here we investigate its interactome network with HT-Y2H.

As Y2H baits, we selected a set of 3024 worm predicted proteins that relate directly or indirectly to multicellular functions (7). Gateway-cloned open reading frames (ORFs) were available in the C. elegans ORFeome 1.1 (8) for 1978 of these selected proteins. Of these, 81 autoactivated the Y2H GAL1::HIS3 reporter gene as Gal4 DNA binding domain fusions (DB-X), and 24 others conferred toxicity to yeast cells. The remaining 1873 baits were screened against two different Gal4 activation domain libraries (AD-wrmcDNA and

References and Notes

10. P. M. Harrison et al., Genome Res. 12, 272 (2002).
16. Materials and methods are available as supporting material on Science Online.
18. Z. Zhang, P. Harrison, M. Gerstein, Genome Res. 12, 1466 (2002).
To estimate the coverage of the HT-Y2H data sets, we manually searched the baits screened here for known interactors in WormPD (10). This search gave rise to 108 interactions, referred to as the “literature” data set (table S1). The Core and Non-Core data sets recapitulated eight and two interactions in this benchmark data set, respectively. Thus, our overall rate of coverage for the First-Pass data set is ~10% ([8 + 2]/108).

To evaluate the accuracy of the HT-Y2H data sets, we reasoned that interactions detected in two different binding assays are unlikely to be experimental false-positives. A representative sample of Y2H interaction pairs from each of these three subsets (33 for Core-1, 62 for Core-2, and 48 for Non-Core) was randomly selected, and tested in a coaffinity purification (co-AP) glutathione S-transferase (GST) pull-down assay (Fig. 1). Bait and prey ORFs were transiently transfected into 293T cells as GST-bait and Myc-prey fusions, and prey ORFs were transiently transfected into baculovirus vectors. Immunoprecipitation using anti-Myc antibodies was performed, and GST-bait protein complexes were analyzed by SDS-PAGE and Coomassie Blue staining. ORF pairs were identified in this co-AP assay if they coimmunoprecipitated with an active GST-bait fusion protein.

In addition to experimental screens, we also performed in silico searches for potentially conserved interactions, or “interologs,” whose orthologous pairs are known to interact in one or more other species (9, 11). Starting from a high-confidence yeast interaction data set (7), reciprocal best-hit BLAST searches (E-value ≤ 10−4) were performed against the worm predicted proteome. In all, 949 potential worm interologs were identified, constituting the interologs data set (7). In addition, the Y2H interactome maps that have been previously generated for individual biological processes (including vulval development, protein degradation, DNA damage response, and germine formation) (9, 12–14) were pooled to define the “scaffold” data set. The HT-Y2H, literature, interologs, and scaffold data sets were combined into Worm Interactome version 5 (WIS), containing 5534 interactions and connecting 15% of the C. elegans proteome (table S1). WIS gives rise to a giant network component of 2898 nodes connected by 5460 edges (Fig. 2A). Similar to other biological networks (15), the worm interactome network exhibits small-world and scale-free properties (Fig. 2B) (7). This data set also allowed us to analyze whether or not evolutionary recent proteins tend to preferentially interact with each other rather than with ancient proteins. We subdivided the nodes of the network into three classes: 748 proteins with a clear ortholog in yeast (“ancient”), 1314 proteins with a clear ortholog in Drosophila, Arabidopsis, or humans but not in yeast (“multicellular”), and 836 proteins with no detectable ortholog outside of C. elegans (“worm”) (7). These three groups seem to connect equally well with each other (Fig. 2C), which suggests that new cellular functions rely on a combination of evolutionarily new and ancient elements, consonant with the classic proposal of evolution as a tinkerer that modifies and adds to pre-existing structures to create new ones (16). Previous studies have related interactome data with genome-wide expression (transcrip-
tome) and phenotypic profiling (phenome) data in *S. cerevisiae* (17). To investigate to what extent different functional genomic assays should correlate in the context of a multicellular organism, we overlapped WI5 with *C. elegans* transcriptome and phenome data sets. Based on a *C. elegans* transcriptome compendium data set (18), we calculated Pearson correlation coefficients (PCCs) for gene pairs involved in Y2H interactions and compared them with randomized data sets (Fig. 2D).

About 150 Core interactions (9.5%) correspond to gene pairs with significantly higher PCCs than expected from random (*P* < 0.05) (table S3). Thus, those pairs can be considered “more biologically likely” because two completely independent approaches point to a functional relationship between the corresponding genes. The remaining pairs are labeled “without additional evidence.” Indeed, it is important to note that lack of coexpression does not suggest that the corresponding interactions are irrelevant. Indeed, 75% of literature pairs, defined as biologically relevant, do not correlate with transcriptome data (Fig. 2D).

We also systematically examined Y2H interactions where both proteins belong to common *C. elegans* expression clusters, or “Topomap mountains” (18). As an example, a highly connected subnetwork derived from mountain 29 (Fig. 2E) contains seven proteins (ABU-1, ABU-8, ABU-11, PQN-5, PQN-54, PQN-57, and PQN-71) that share common domains (DUF139 domain and cysteine-rich repeat). Furthermore, these proteins are all expressed in the pharynx (19–21), which suggests that they may act together in pharynx function or development.

For relatively small-scale *S. cerevisiae* and *C. elegans* interactome data sets, physical interactions pointed to genes that share similar phenotypes when knocked out or knocked down (17). To evaluate this idea for the *C. elegans* interactome, we assembled a collection of phenotypic data based on RNA interference (RNAi) knockdown experiments from WormBase (7, 22), and we calculated the percentage of protein interaction pairs that share embryonic lethal phenotypes for the interaction data sets and their randomized
controls and found a twofold enrichment for the Core and First-Pass data sets (Fig. 2F). Similar correlations were also observed for the maternal sterile phenotype and four groups of postembryonic phenotypes (23). Because protein-protein interactions for which both genes are coexpressed across many conditions and show similar phenotype(s) when knocked down should be considered particularly likely, the global correlation described above illustrate how biological hypotheses can be derived from overlapping interactome, transcriptome, and phenome data sets (table S3).

In S. cerevisiae, two proteins that have many interaction partners in common are more likely to be related biologically (24). We examined the C. elegans interactome network for the presence of highly connected neighborhoods by determining the mutual clustering coefficient between proteins in the network (table S4) (24). As an example, we examined the properties of one of the clusters containing such a high-scoring protein pair: VAB-3/C49A1.4 (Fig. 3). VAB-3 and C49A1.4 have strong similarity to the products of the Drosophila genes eyeless (ey) and eyes absent (eya), respectively, but not to each other. EY and EYA are components of a conserved network of transcription factors that regulate eye development (25).

VAB-3 and C49A1.4 are part of a highly interconnected subnetwork in W15 (Fig. 3) with proteins that are known or suspected to be functionally linked to VAB-3 and C49A1.4, or to their respective orthologs in other organisms. These include (i) EGL-27, which negatively regulates MAB-5 in hermaphrodites (26) and is linked to MAB-5 through C49A1.4; (ii) WRT-2, an interactor of C49A1.4 with similarity to Drosophila Hedgehog, which alleviates repression of eya expression by Cubitus interruptus (27); and (iii) CEH-33 and CEH-35, two of four members of the sine oculis homeobox gene family, which is involved in the same Drosophila regulatory network of transcription factors as ey and eya (28). Finally, eight proteins in this cluster are annotated in WormPD as involved in membrane function, which suggests a functional relationship between the eyeless transcription network and membrane activity.

Together with interologs and previously described interactions, the Y2H data set provides functional hypotheses for thousands of uncharacterized proteins in the C. elegans proteome. Integration with other functional genomic data indicates that the correlation between transcriptome and interactome data, although significant, is lower than what would be expected from observations made in yeast (17). This observation applies to both transcriptome and interactome data, which both genes are coexpressed across many conditions and show similar phenotypes when knocked down should be expected from observations made in yeast (17).

Our current interactome map also illustrates how a human interactome project would benefit from an ORFeome cloning project using recombinational cloning systems, such as Gateway (8). Indeed, recombinationally cloned ORFs can be shuffled at will into various expression vectors needed for different types of protein interaction assays, as exemplified by our ability to transfer bait- and prey-encoding ORFs into Myc- and GST-tagged vectors to validate Y2H interactions.

References and Notes
7. See supporting material on Science Online.
11. L. R. Matthews et al., Genome Res. 11, 2120 (2001).
23. H. Ge, unpublished observations.
29. We thank members of M.V.’s laboratory for their support and input and help; C. Boone, G. Achaz and D. Allinger for discussions; the sequencing staff at Agencourt Biosciences for technical assistance; the ORFeome meeting participants for their input; C. McCowan, T. Cling-ingham, and C. You for administrative assistance; and C. Fraughton for laboratory support. This work was supported by a grant from NIH and NIGMS awarded to M.V. Other support includes an NSF award (K.C.G.); NIGMS grants (S.v.d.H., S.E.M., F.P.R., G.F.B); and Fellowships from EMBO (P.-O.V.), NSF (D.S.G), Ryan, Milton (L.V.Z.), and Leukemia Research Foundation (C.S., A.C.); an institutional HHMI grant (F.P.R., G.F.B); and Fellowships from EMBO (P.-O.V.), NSF (D.S.G), Ryan, Milton (L.V.Z.), and Leukemia Research Foundation (C.S., A.C.).

Material and Methods
Supporting Online Material
www.sciencemag.org/cgi/content/full/1091403/DC1

Table S1 to S5
Fig S1
References
Table S1 to S5
Fig S1
References
Table S1 to S5
Fig S1
References
11 September 2003; accepted 1 December 2003
Published online 2 January 2004.
10.1126/science.1091403
Include this information when citing this paper.