LETTERS

Predictive models of molecular machines involved in *Caenorhabditis elegans* early embryogenesis

Kristin C. Gunsalus[1]*, Hui Ge[2]*, Aaron J. Schetter[1]*, Debra S. Goldberg[3]*, Jing-Dong J. Han[2], Tong Hao[2], Gabriel F. Berriz[3], Nicolas Bertin[2], Jerry Huang[1], Ling-Shiang Chuang[1], Ning Li[2], Ramamurthy Mani[3], Anthony A. Hyman[4], Birte Sönnichsen[5], Christophe J. Echeverri[5], Frederick P. Roth[3], Marc Vidal[2] & Fabio Piano[1]

Although numerous fundamental aspects of development have been uncovered through the study of individual genes and proteins, system-level models are still missing for most developmental processes. The first two cell divisions of *Caenorhabditis elegans* embryogenesis constitute an ideal test bed for a system-level approach. Early embryogenesis, including processes such as cell division and establishment of cellular polarity, is readily amenable to large-scale functional analysis. A first step toward a system-level understanding is to provide 'first-draft' models both of the molecular assemblies involved[1] and of the functional connections between them. Here we show that such models can be derived from an integrated gene/protein network generated from three different types of functional relationship[2]: protein interaction[3], expression profiling similarity[4] and phenotypic profiling similarity[5], as estimated from detailed early embryonic RNA interference phenotypes systematically recorded for hundreds of early embryogenesis genes[6]. The topology of the integrated network suggests that *C. elegans* early embryogenesis is achieved through coordination of a limited set of molecular machines. We assessed the overall predictive value of such molecular machine models by dynamic localization of ten previously uncharacterized proteins within the living embryo.

Global correlations between transcriptome profiling and interactome data sets have been used to derive network graphs that combine similarity relationships from transcription profiling with physical interactions between proteins[3,7–13]. Suggestive correlations between interactome or transcriptome data and phenotypic data sets[5,10,14,15] support the notion that these three types of data might complement one another in predicting functional relationships.

To model *C. elegans* early embryogenesis globally, we generated network graphs in which each node represents an early embryogenesis gene[6] and its product(s), and each edge represents a potential functional connection based on one of three data sets (Fig. 1a): (1) 6,572 binary physical interactions between 3,848 *C. elegans* proteins (WI7 data set; Supplementary Methods and Supplementary Table S1)[3]; (2) expression profiling similarity above a given threshold (transcriptional Pearson correlation coefficients (transcriptional PCCs) from a compendium of *C. elegans* microarray profiles[4]); and (3) phenotypic similarity above another threshold (described below).

For each of the 661 early embryogenesis genes identified[6], we used an RNA interference (RNAi) phenotypic signature[5] consisting of a vector describing specific cellular defects in early embryogenesis[6]. We defined a measure of phenotypic similarity between early embryogenesis genes as the uncentred Pearson correlation coefficient (phenotypic PCC) for each pair of signatures (Supplementary Methods). The level of phenotypic similarity correlates with similar functional attributes (Fig. 1b, top right), as measured by shared Gene Ontology (GO) terms (which provide a controlled vocabulary for gene function)[16]. We performed hierarchical clustering to group genes by phenotypic similarity, and observed that clusters tend to show significant enrichment for specific gene functions (Fig. 1b; see also Supplementary Methods, Supplementary Fig. S1 and Supplementary Table S2). These results suggest that phenotypic PCC derived from RNAi data represents a reasonable way to compare phenotypes quantitatively.

To evaluate whether all three functional relationships (physical interaction, expression similarity and phenotypic similarity) can be merged into predictive models, we asked whether they show correlations among early embryogenesis genes/proteins (Fig. 2). First, we found that the products of early embryogenesis genes are more interconnected by direct protein interactions than expected by chance (Fig. 2a; see also Supplementary Methods and Supplementary Table S3). Second, compared to genes partitioned randomly (Fig. 2b, right), genes clustered by phenotypic similarity encode proteins that are more likely to interact physically with one another, either directly (Fig. 2b, top left) or indirectly through a single shared interactor (Fig. 2b, bottom left). Third, expression correlations are significantly higher between protein interactors, early embryogenesis gene pairs and pairs in the same phenocluster (mean transcriptional PCC = 0.17, 0.16 and 0.19, respectively) relative to gene pairs selected at random from the genome (mean transcriptional PCC = 0.05) (Fig. 2c). An even more notable increase in expression correlation (ten times the level between random pairs) is seen among early embryogenesis or intra-phenocluster pairs that also interact directly (Fig. 2c).

Consistent with these global correlations, we found that the proportion of direct interactors rises as a function of both phenotype correlation (Fig. 2d) and expression correlation (Fig. 2e). Similarly, there is a positive relationship between phenotype and expression correlation values, particularly when only direct protein interactions are considered (Fig. 2f). Notably, interactors with strong expression correlation show strong phenotype correlation, suggesting membership in a common molecular assembly; interactors with low correlations of both types may represent either false positives or a different relationship (for example, a regulatory interaction). Overall, phenotypic and expression correlations both show a strong inverse relationship with distance in the interactome network (Fig. 2g). On the basis of the above correlations, we assigned edges between pairs of nodes

[1]Center for Comparative Functional Genomics, Department of Biology, New York University, New York, New York 10003, USA. [2]Center for Cancer Systems Biology and Department of Cancer Biology, Dana-Farber Cancer Institute and Department of Genetics, Harvard Medical School, Boston, Massachusetts 02115, USA. [3]Department of Biological Chemistry and Molecular Pharmacology, Harvard Medical School, Massachusetts 02115, USA. [4]Max Planck Institute of Molecular Cell Biology and Genetics, 01307 Dresden, Germany. [5]Cenix BioScience GmbH, 01307 Dresden, Germany.
*These authors contributed equally to this work.

based on expression similarity (transcriptional PCC ≥ 0.7) or phenotypic similarity (phenotypic PCC ≥ 0.5).

The integrated early embryogenesis network—joining all 661 early embryogenesis genes/proteins by the union of all three types of relationship (Fig. 3a)—contains a main component with 31,173 edges characterized by an average of 0.9, 5.0 and 44 edges per node for protein interaction (Int), expression similarity (Tr) and phenotypic similarity (Ph), respectively. In this network, the number of gene/protein pairs with doubly supported edges is significantly higher than expected by chance ($P = 10^{-34}$, 10^{-61} and 10^{-109} for Int–Tr, Ph–Tr and Int–Ph associations, respectively). We examined the portion of this integrated early embryogenesis network containing only links with two or more types of functional support (Fig. 3b). In contrast to the full network, the topology of this 'multiple support network'—which contains about half (305) of the early embryogenesis genes/proteins—reveals distinct groups of highly interconnected genes/proteins and few or no links between the groups. To assess the predictive value of the early embryogenesis network on a global scale, we analysed the individual and combined networks for their ability to predict a specific shared function between two linked gene pairs using GO annotations (Supplementary Methods). Each individual network has significant power to detect shared function between linked gene pairs, and combining data types generally results in

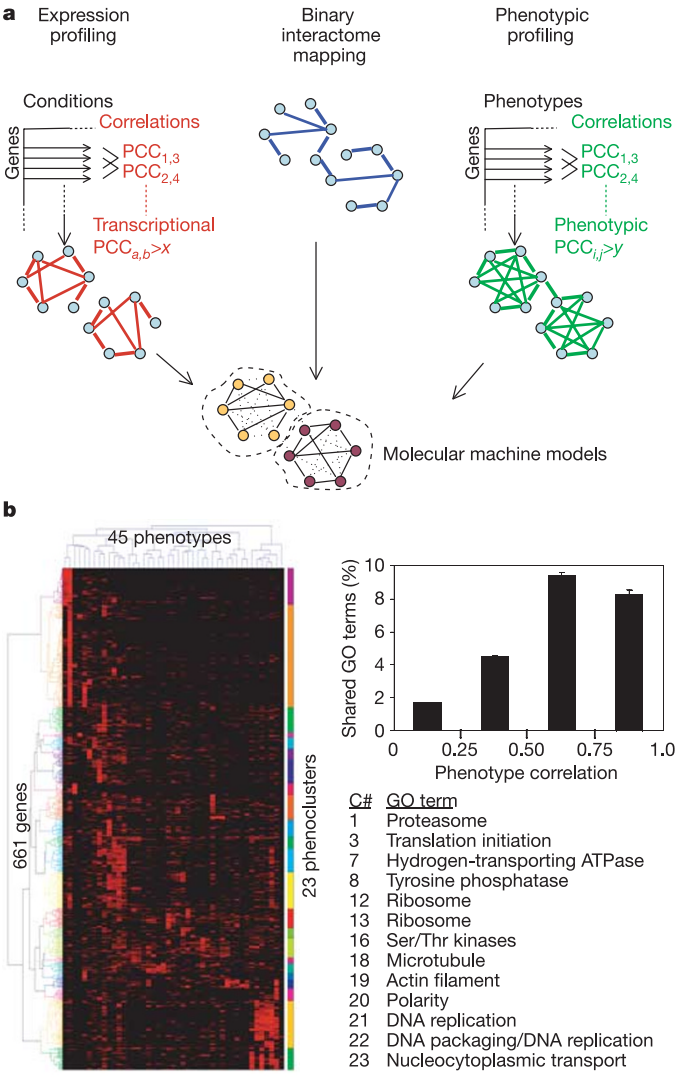

Figure 1 | Integrated networks and phenotypic profiling. **a**, Pairwise relationships between genes/proteins are determined from correlated transcript abundance (transcriptional PCC) (left), physical protein–protein interactions (centre) and phenotypic correlation (phenotypic PCC) (right). Graphs represent genes/proteins as nodes and relationships (transcriptional PCCs above threshold x, physical interactions, and phenotypic PCCs above threshold y) as edges. Highly interconnected regions represent models of molecular machines or processes (bottom). **b**, Six-hundred and sixty-one early embryogenesis genes clustered by phenotypic similarity using high-content early embryogenesis phenotypes (left; see also Supplementary Fig. S1). The fraction of shared functional annotations increases with phenotypic similarity (top right); most phenoclusters (indicated by cluster number, C#) are enriched for specific GO functional annotations (bottom right; see also Supplementary Table S1). Error bars represent standard errors of the mean.

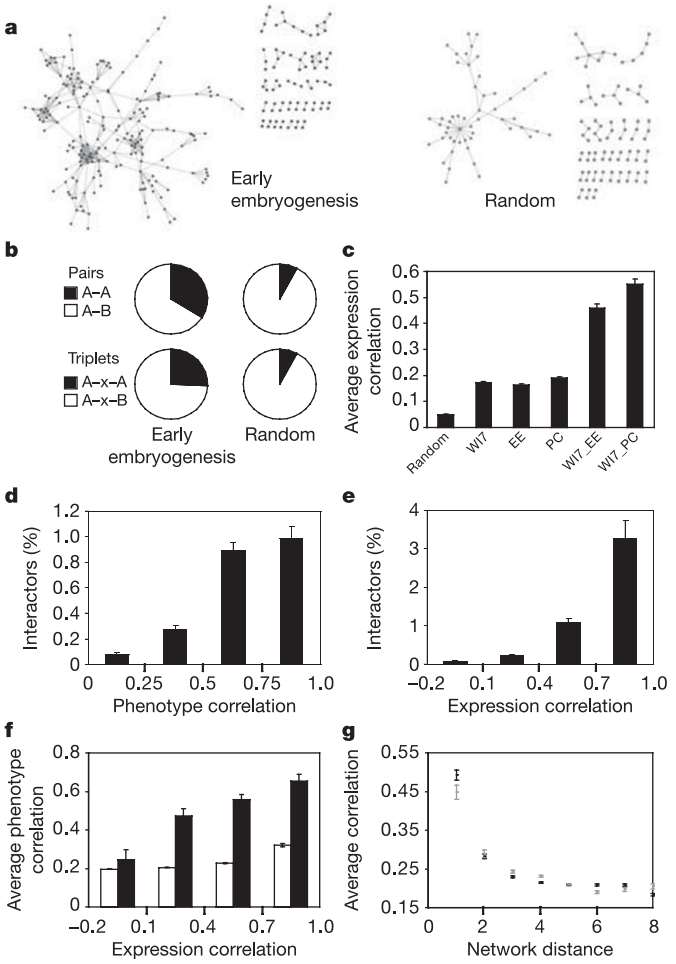

Figure 2 | Correlations between data sets (see also Supplementary Fig. S2). **a**, The early embryogenesis interactome subnetwork from WI7 (left) exhibits higher connectivity than networks of proteins chosen randomly (example at right; see also Supplementary Table S3). **b**, The early embryogenesis interactome subnetwork is enriched for interactions within the same phenocluster (A–A and A–x–A) relative to interactions between phenoclusters (A–B and A–x–B). **c**, Interacting proteins (WI7), random early embryogenesis (EE) pairs, intra-phenocluster early embryogenesis pairs (PC), pairs of interacting early embryogenesis proteins (WI7_EE) and interacting early embryogenesis proteins from common phenoclusters (WI7_PC) all show higher expression correlation than random pairs. **d**, **e**, The proportion of physical interactions increases with phenotypic (**d**) and expression correlation (**e**). **f**, Early embryogenesis genes with similar expression profiles are more likely to share similar RNAi phenotypes. All early embryogenesis gene pairs (open bars) and interacting early embryogenesis proteins (filled bars) were binned by expression correlation and plotted against average phenotypic correlation. **g**, Phenotype and expression correlation increase with interactome proximity. Average phenotype (black) and expression (grey) correlation decrease for early embryogenesis protein pairs as their distance (shortest path) increases. Error bars in **c**–**g** represent standard errors of the mean.

higher accuracy but lower sensitivity (Supplementary Tables S4 and S5). Considering only pairs for which both members have some GO annotation, the accuracy of the multiple support network is very high, with 88% of the links sharing a specific functional annotation.

To generate models representing the higher-level organization underlying early embryogenesis, we identified densely interconnected regions in the multiple support network using a graph theoretic clustering algorithm[17] followed by manual refinement based on functional annotations (Fig. 3b; see also Supplementary Table S6). We distinguished two types of highly interconnected regions among the resulting models. The first type of model contains a high density of links supported by both protein interactions and phenotypic correlations. These models represent known molecular complexes that constitute discrete molecular machines within the cell, such as the ribosome, proteasome, mitochondrial F_1F_0 ATPase, vacuolar H^+ ATPase, anaphase-promoting complex (APC) and COPI coatomer, as well as complexes involved in translation initiation, nucleocytoplasmic transport and cell polarity. Virtually all of the edges in the graph that are supported by all three types of evidence (41 out of 43 edges between 50 nodes) fall into such complexes. Proteins within such complexes function together as one physical unit, and depletion of any single member is likely to result in a very similar phenotypic profile.

The second type of model is dominated by edges supported by both phenotypic and expression correlations, containing few physical interactions. These models harbour genes that participate in distinct yet functionally interdependent cellular processes. Examples include messenger RNA/protein metabolism (mRNA transcription and processing, translational control, and protein modification and trafficking), chromosome maintenance/nucleocytoplasmic transport (DNA replication licensing and synthesis, chromosome segregation, nucleoporins and importins), and oocyte integrity/meiosis (oocyte development, extra-embryonic matrix and eggshell formation, and regulation of meiotic events). Within these models smaller molecular machines are found, supported by physical interactions and phenotypic similarity, such as the translation initiation, COPI coatomer, DNA replication licensing and importin complexes. Because current interactome maps have sampled only a small fraction of true interactions[3,18], such coordinated process models may serve to predict undiscovered protein interactions. Alternatively, these models may represent a qualitatively different type of functional unit, in which the phenotypic and expression profiling links reflect functional interdependencies dictated by the logical structure of the network, while the few protein interactions represent the physical path of information flow.

Putative functional interdependency, or cross-talk, between cellular events is evidenced in the multiple support network by links connecting distinct cellular processes. For instance, multiple edges

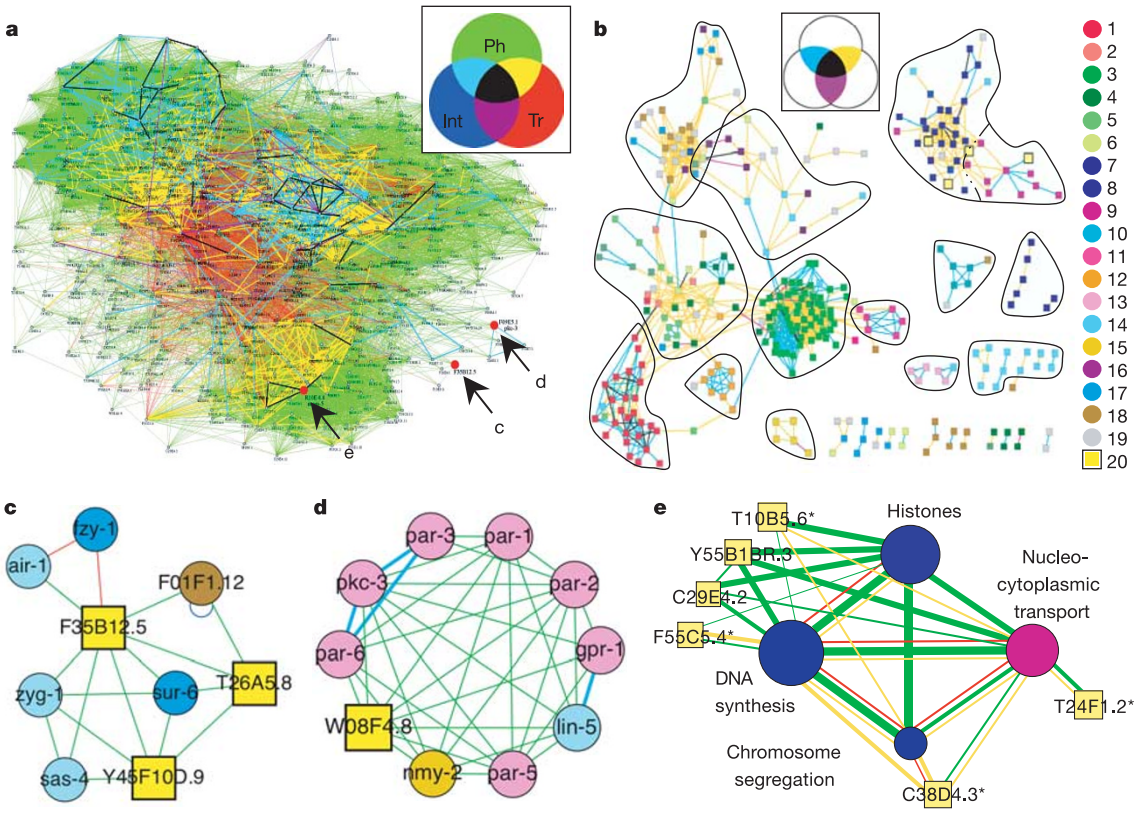

Figure 3 | Integrated network analysis. a, Entire early embryogenesis network graph (see also the Supplementary Data files). A Venn diagram (inset) shows the colour system for labelling edges based on available evidence: phenotypic profiling similarity (Phe; green), expression profiling similarity (Tr; red), physical interaction (Int; blue) and overlapping combinations of data types (intersecting regions). **b**, Multiple support network containing 305 nodes joined by 1,036 edges, each supported by two or three types of functional evidence. Predicted molecular machines are encircled (Supplementary Table S6 provides details). Nodes are colour-coded by function: 1, proteasome; 2, protein degradation; 3, ribosome/protein synthesis; 4, translational control; 5, protein/vesicular trafficking; 6, RNA synthesis/processing/binding; 7, histone; 8, DNA synthesis/replication and chromosome segregation; 9, nucleocytoplasmic transport; 10, APC; 11, mitochondrial F_1F_0 ATPase; 12, vacuolar H^+ ATPase; 13, cell polarity; 14, microtubule cytoskeleton; 15, actin cytoskeleton; 16, cell cycle; 17, signal transduction; 18, metabolism; 19, other/unknown; 20, analysed by protein localization. **c**–**e**, Subnetworks with proteins of unknown function (yellow nodes) analysed by localization (Fig. 4). **c**, Centrosome model. **d**, PAR cell polarity model. **e**, Nuclear function model. Fifty-five genes/proteins in four functional categories were grouped into 'metanodes' (with each metanode representing a collection of individual nodes). Line weights and metanode sizes approximate the number of underlying individual links and nodes, respectively. Four unknowns in this model are present in the multiple support network (asterisks in **e**).

are observed between DNA replication, DNA synthesis and nuclear membrane functions, and between regulation of protein synthesis, protein and vesicular trafficking, and protein degradation. We also observe connections and some overlap between genes required for oocyte integrity (oocyte development, extra-embryonic matrix and eggshell formation) and regulation of meiosis.

To investigate predictions from the early embryogenesis models, we selected ten genes of unknown function linked to groups of functionally characterized genes. To analyse potential participation in a molecular machine using an assay that was independent of the data used to generate the models, we used green fluorescent protein (GFP)-tagged proteins and visualized *in vivo* dynamic subcellular localization during early embryogenesis (Fig. 4; see also Supplementary Movies). We tested proteins with connections to three different early embryogenesis models: centrosomal function (Fig. 3c), cell polarity (Fig. 3d) and an elaborate network of molecular assemblies involved in DNA replication, chromatin architecture and nucleocytoplasmic transport (Fig. 3e).

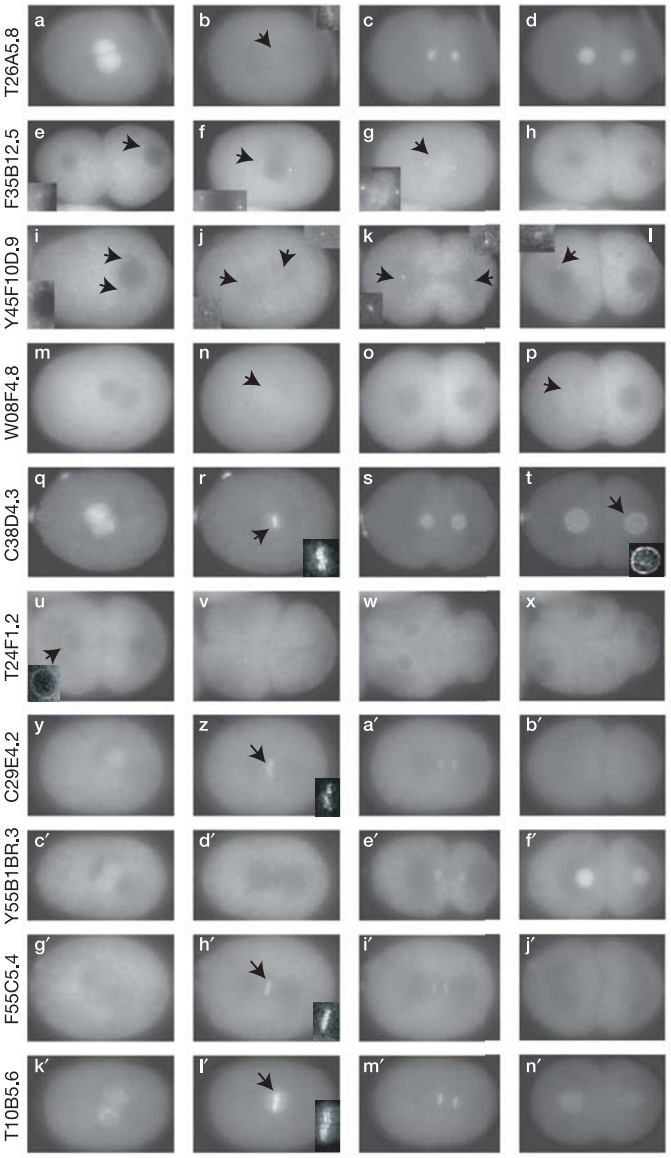

Figure 4 | Embryonic localization patterns of GFP-tagged fusion proteins analysed by time-lapse microscopy are largely consistent with model predictions. Arrows indicate areas expanded in insets (except in **n**, **p**). **a**–**d**, GFP–T26A5.8: nuclear during interphase and chromosomal during mitosis. **e**–**h**, GFP–F35B12.5 (SAS-5): appears on centriole and spindle. **i**–**l**, GFP–Y45F10D.9 (SAS-6): centriolar. **m**–**p**, GFP–W08F4.8: weakly nuclear (arrows in **n**, **p**) during mitosis but excluded from nucleus during interphase (compare two cells in **p**). **q**–**t**, GFP–C38D4.3: shuttles between nuclear periphery during interphase onto chromosomes during mitosis (double line at metaphase suggests kinetochore localization). **u**–**x**, GFP–T24F1.2: enriched on nuclear membrane. **y**–**b′**, GFP–C29E4.2: transient chromosomal localization (arrow in **z**). **c′**–**f′**, GFP–Y55B1BR.3: nuclear signal peaking during S phase (**f′**). **g′**–**n′**, GFP–F55C5.4 (**g′**–**j′**) and GFP–T10B5.6 (KNL-3) (**k′**–**n′**): transient chromosomal localization peaking at metaphase (double-line pattern suggests kinetochore localization). See Supplementary Movies for corresponding time-lapse recordings.

The centrosomal function model (Fig. 3c) connects several centriolar proteins (ZYG-1, AIR-1, SAS-4)[19–22] and FZY-1, a *C. elegans* homologue of Cdc20, which regulates the APC and is required for proper chromosome separation[23]. Remarkably, two of the three proteins we targeted (F35B12.5 and Y45F10D.9) localized to a pair of perinuclear puncta with dynamic behaviour matching that of centrioles (Fig. 4e–l; see also Supplementary Movies S2 and S3). These dynamics have subsequently been confirmed independently[24–26]. The third protein (T26A5.8) instead localized to the nucleus and faintly to metaphase chromosomes (Fig. 4a–d; see also Supplementary Movie S1); the nature of its apparent connection to centrosomal function remains to be elucidated.

The cell polarity model (Fig. 3d) contains all of the PAR proteins (partitioning-defective regulators of polarity) present in the early embryogenesis data set, as well as additional proteins known to be important for establishing polarity in *C. elegans*[27]. Interestingly, links in this subnetwork resemble mass spectrometry results for mammalian PAR homologues[28], and both implicate a homologue of *Saccharomyces cerevisiae* Cdc37 (W08F4.8 in Fig. 3d) in polarity. GFP–W08F4.8 did not reveal any asymmetric localization (a prominent feature of several PAR proteins), but appeared to fill the nucleus at prophase and be excluded from the nucleus at telophase (Fig. 4m–p; see also Supplementary Movie S4). The mammalian homologue of W08F4.8 (Cdc37) binds to the PAR-4 homologue (tumour suppressor LKB1), which shuttles in and out of the nucleus to regulate polarity[29]. The localization pattern of W08F4.8 suggests that it may participate in the nuclear shuttling of PAR-4 (PAR-4 is not shown because it is not in the early embryogenesis data set).

In each of six tests for the nuclear function model (Fig. 3e), the localization pattern of the targeted protein was consistent with its predicted role. The first, C38D4.3, is linked by expression and phenotypic similarity to members of the nuclear pore complex (*npp-2* and *npp-10*), a homologue of the centromere protein CENP-C (*hcp-4*) and DNA synthesis components (*mcm-2*, DNA polymerase B and a DNA topoisomerase II homologue). GFP–C38D4.3 localized in a remarkable dynamic pattern (Fig. 4q–t; see also Supplementary Movie S5), shuttling between the nuclear membrane (during interphase) and the chromosomes (during mitosis), appearing to coalesce at the nuclear envelope onto chromosomes (Fig. 4q), and then during metaphase forming a double-line pattern reminiscent of centromere/kinetochore proteins (Fig. 4r). The second protein, T24F1.2, is linked to several nuclear pore complex components by phenotypic similarity and to *ran-1*, a nuclear transport control factor, via expression correlation. Consistent with these connections, GFP–T24F1.2 localized to the nuclear envelope throughout the cell cycle (Fig. 4u–x; see also Supplementary Movie S6).

The last four genes/proteins tested have multiple links to the DNA synthesis and histone subnetworks (Fig. 3e). One of these, F55C5.4, is linked by both expression and phenotype to several DNA replication licensing factors (*mcm-3*, *mcm-5*, *mcm-6*, *mcm-7*) and DNA synthesis components (*lig-1*, *rnr-2*). GFP-tagged C29E4.2 and F55C5.4 both localized to condensing chromosomes exclusively around metaphase (Fig. 4g′–j′ and y–b′, respectively; see also Supplementary Movies S7 and S9). In contrast, GFP–Y55B1BR.3 showed a diffuse nuclear localization that peaked during S phase (Fig. 4c′–f′; see also Supplementary Movie S8). Finally, T10B5.6 is linked by phenotype and expression to both *hcp-4* (CENP-C) and *npp-19* (a nucleoporin),

and the GFP-tagged protein showed diffuse nuclear localization alternating with a double-line pattern typical of kinetochore/centromere proteins (Fig. 4k′–n′; see also Supplementary Movie S10). This protein's dynamic localization pattern and its function at the kinetochore was recently confirmed independently[30]. In total, eight out of ten experimental tests for potential new components of molecular machines gave rise to supportive evidence. Two localization experiments remain inconclusive, although for one (W08F4.8), proteomic experiments[28] support the hypothesis formulated here.

We have shown that phenotypic profiling data can be combined with interactome and expression profiling data to generate a network of functional relationships for *C. elegans* early embryogenesis. The combined evidence network suggests that the molecular machines acting in early embryogenesis are highly interconnected, and are likely to operate together through regulatory molecules that coordinate their activities. We have tested ten predictions from this integrated network by observing the dynamic *in vivo* localization pattern of GFP fusion proteins. The integrated network is a potential reservoir for hundreds of testable predictions about cellular processes in the early embryo. The approach presented here is scalable and can be extended to include additional data types. This general strategy is applicable to other biological processes and other organisms, including humans.

METHODS

Clustering and network analysis. Agglomerative hierarchical clustering of genes based on phenotypic similarity, construction of network graphs from protein–protein interaction data and similarity in expression and phenotypic profiles and corresponding statistical analyses are described in Supplementary Methods. GO annotations from WormBase version WS100 were used to test phenoclusters for functional enrichment and to test the integrated early embryogenesis network for its ability to predict a shared function between two linked gene/protein pairs, as detailed in Supplementary Methods.

Localization studies. For each gene selected, the sequence between the predicted initiation and termination codons was cloned into GFP vectors driving expression in the germ line and soma. Transgenic animals expressing extrachromosomal arrays were generated by injection, and animals from the F_2 generation were assayed for localization. Microscopy was carried out on Leica DMLA or DMRA microscopes using × 100 (1.3 N.A.) objectives and GFP filters. (See Supplementary Methods for details.)

Received 30 March; accepted 23 May 2005.

1. Alberts, B. The cell as a collection of protein machines: preparing the next generation of molecular biologists. *Cell* **92,** 291–294 (1998).
2. Vidal, M. A biological atlas of functional maps. *Cell* **104,** 333–339 (2001).
3. Li, S. *et al.* A map of the interactome network of the metazoan *C. elegans*. *Science* **303,** 540–543 (2004).
4. Kim, S. K. *et al.* A gene expression map for *Caenorhabditis elegans*. *Science* **293,** 2087–2092 (2001).
5. Piano, F. *et al.* Gene clustering based on RNAi phenotypes of ovary-enriched genes in *C. elegans*. *Curr. Biol.* **12,** 1959–1964 (2002).
6. Sönnichsen, B. *et al.* Full-genome RNAi profiling of early embryogenesis in *Caenorhabditis elegans*. *Nature* **434,** 462–469 (2005).
7. Marcotte, E. M., Pellegrini, M., Thompson, M. J., Yeates, T. O. & Eisenberg, D. A combined algorithm for genome-wide prediction of protein function. *Nature* **402,** 83–86 (1999).
8. Ge, H., Liu, Z., Church, G. M. & Vidal, M. Correlation between transcriptome and interactome mapping data from *Saccharomyces cerevisiae*. *Nature Genet.* **29,** 482–486 (2001).
9. Jansen, R., Greenbaum, D. & Gerstein, M. Relating whole-genome expression data with protein-protein interactions. *Genome Res.* **12,** 37–46 (2002).
10. Walhout, A. J. *et al.* Integrating interactome, phenome, and transcriptome mapping data for the *C. elegans* germline. *Curr. Biol.* **12,** 1952–1958 (2002).
11. Lee, I., Date, S. V., Adai, A. T. & Marcotte, E. M. A probabilistic functional network of yeast genes. *Science* **306,** 1555–1558 (2004).
12. Bader, J. S., Chaudhuri, A., Rothberg, J. M. & Chant, J. Gaining confidence in high-throughput protein interaction networks. *Nature Biotechnol.* **22,** 78–85 (2004).
13. Jansen, R. *et al.* A Bayesian networks approach for predicting protein-protein interactions from genomic data. *Science* **302,** 449–453 (2003).
14. Boulton, S. J. *et al.* Combined functional genomic maps of the *C. elegans* DNA damage response. *Science* **295,** 127–131 (2002).
15. Begley, T. J., Rosenbach, A. S., Ideker, T. & Samson, L. D. Damage recovery pathways in *Saccharomyces cerevisiae* revealed by genomic phenotyping and interactome mapping. *Mol. Cancer Res.* **1,** 103–112 (2002).
16. Ashburner, M. *et al.* Gene ontology: tool for the unification of biology. The Gene Ontology Consortium. *Nature Genet.* **25,** 25–29 (2000).
17. Bader, G. D. & Hogue, C. W. An automated method for finding molecular complexes in large protein interaction networks. *BMC Bioinformatics* **4,** 2 (2003).
18. Giot, L. *et al.* A protein interaction map of *Drosophila melanogaster*. *Science* **302,** 1727–1736 (2003).
19. O'Connell, K. F. *et al.* The *C. elegans zyg-1* gene encodes a regulator of centrosome duplication with distinct maternal and paternal roles in the embryo. *Cell* **105,** 547–558 (2001).
20. Schumacher, J. M., Ashcroft, N., Donovan, P. J. & Golden, A. A highly conserved centrosomal kinase, AIR-1, is required for accurate cell cycle progression and segregation of developmental factors in *Caenorhabditis elegans* embryos. *Development* **125,** 4391–4402 (1998).
21. Kirkham, M., Muller-Reichert, T., Oegema, K., Grill, S. & Hyman, A. A. SAS-4 is a *C. elegans* centriolar protein that controls centrosome size. *Cell* **112,** 575–587 (2003).
22. Leidel, S. & Gönczy, P. SAS-4 is essential for centrosome duplication in *C. elegans* and is recruited to daughter centrioles once per cell cycle. *Dev. Cell* **4,** 431–439 (2003).
23. Kitagawa, R., Law, E., Tang, L. & Rose, A. M. The Cdc20 homolog, FZY-1, and its interacting protein, IFY-1, are required for proper chromosome segregation in *Caenorhabditis elegans*. *Curr. Biol.* **12,** 2118–2123 (2002).
24. Delattre, M. *et al.* Centriolar SAS-5 is required for centrosome duplication in *C. elegans*. *Nature Cell Biol.* **6,** 656–664 (2004).
25. Dammermann, A. *et al.* Centriole assembly requires both centriolar and pericentriolar material proteins. *Dev. Cell* **7,** 815–829 (2004).
26. Leidel, S., Delattre, M., Cerutti, L., Baumer, K. & Gönczy, P. SAS-6 defines a protein family required for centrosome duplication in *C. elegans* and in human cells. *Nature Cell Biol.* **7,** 115–125 (2005).
27. Schneider, S. Q. & Bowerman, B. Cell polarity and the cytoskeleton in the *Caenorhabditis elegans* zygote. *Annu. Rev. Genet.* **37,** 221–249 (2003).
28. Brajenovic, M., Joberty, G., Kuster, B., Bouwmeester, T. & Drewes, G. Comprehensive proteomic analysis of human Par protein complexes reveals an interconnected protein network. *J. Biol. Chem.* **279,** 12804–12811 (2004).
29. Baas, A. F., Smit, L. & Clevers, H. LKB1 tumour suppressor protein: PARtaker in cell polarity. *Trends Cell Biol.* **14,** 312–319 (2004).
30. Cheeseman, I. M. *et al.* A conserved protein network controls assembly of the outer kinetochore and its ability to sustain tension. *Genes Dev.* **18,** 2255–2268 (2004).

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

Acknowledgements This work was supported by grants from NSF (to K.C.G. and D.S.G.), NIH/NHGRI (to M.V. and F.P.R.), NIH/NICHD (to F.P.) and Taplin Funds for Discovery (to F.P.R.). We thank members of our laboratories for discussions and comments on the manuscript.

Author Contributions K.C.G., H.G., D.S.G., J.-D.J.H., T.H., N.B., N.L., J.H., G.F.B. and R.M. performed informatic analyses; A.J.S. and L.-S.C. performed cloning and localization experiments; K.C.G., F.P.R., M.V. and F.P. wrote the manuscript; B.S., C.E. and A.A.H. shared critical data before publication; and F.P.R., M.V. and F.P. provided guidance for experimental and informatic analyses and interpretation of results.

Author Information Reprints and permissions information is available at npg.nature.com/reprintsandpermissions. The authors declare no competing financial interests. Correspondence and requests for materials should be addressed to F.P. (fp1@nyu.edu), M.V. (marc_vidal@dfci.harvard.edu) or F.P.R. (fritz_roth@hms.harvard.edu).