

A High-Resolution *C. elegans* Essential Gene Network Based on Phenotypic Profiling of a Complex Tissue

Rebecca A. Green,¹ Huey-Ling Kao,^{3,10} Anjon Audhya,^{2,10} Swathi Arur,⁴ Jonathan R. Mayers,² Heidi N. Fridolfsson,⁵ Monty Schulman,³ Siegfried Schloissnig,⁷ Sherry Niessen,⁸ Kimberley Laband,¹ Shaohe Wang,¹ Daniel A. Starr,⁵ Anthony A. Hyman,⁷ Tim Schedl,⁶ Arshad Desai,^{1,11} Fabio Piano,^{3,9,11} Kristin C. Gunsalus,^{3,9,*} and Karen Oegema^{1,*}

¹Ludwig Institute for Cancer Research and Department of Cellular and Molecular Medicine, University of California, San Diego, CMM-East 3053, 9500 Gilman Drive, La Jolla, CA 92093, USA

²Department of Biomolecular Chemistry, University of Wisconsin-Madison Medical School, 1300 University Avenue, Madison, WI 53706, USA

³Center for Genomics and Systems Biology, Department of Biology, New York University, 12 Waverly Place, 8th Floor, New York, NY 10003, USA

⁴Department of Genetics, The University of Texas MD Anderson Cancer Center, 1515 Holcombe Blvd, Unit 1010, Houston, TX 77030, USA

⁵Department of Molecular and Cellular Biology, College of Biological Sciences, University of California, Davis, One Shields Avenue, Davis, CA 95616, USA

⁶Department of Genetics, Washington University School of Medicine, Saint Louis, Missouri, 63110 Washington University, St. Louis, MO 63110, USA

⁷Max Planck Institute of Molecular Cell Biology and Genetics, Pfotenhauerstrasse 108, 01307 Dresden, Germany

⁸The Skaggs Institute for Chemical Biology and Department of Chemical Physiology, The Center for Physiological Proteomics, The Scripps Research Institute, La Jolla, CA 92037, USA

⁹New York University Abu Dhabi, Abu Dhabi, United Arab Emirates

¹⁰These authors contributed equally to this work

¹¹These authors contributed equally to this work

*Correspondence: kcg1@nyu.edu (K.C.G.), koegema@ucsd.edu (K.O.)

DOI 10.1016/j.cell.2011.03.037

SUMMARY

High-content screening for gene profiling has generally been limited to single cells. Here, we explore an alternative approach—profiling gene function by analyzing effects of gene knockdowns on the architecture of a complex tissue in a multicellular organism. We profile 554 essential *C. elegans* genes by imaging gonad architecture and scoring 94 phenotypic features. To generate a reference for evaluating methods for network construction, genes were manually partitioned into 102 phenotypic classes, predicting functions for uncharacterized genes across diverse cellular processes. Using this classification as a benchmark, we developed a robust computational method for constructing gene networks from high-content profiles based on a network context-dependent measure that ranks the significance of links between genes. Our analysis reveals that multi-parametric profiling in a complex tissue yields functional maps with a resolution similar to genetic interaction-based profiling in unicellular eukaryotes—pinpointing subunits of macromolecular complexes and components functioning in common cellular processes.

INTRODUCTION

A major challenge of the postgenomic era is to translate the parts lists generated by genome sequencing into maps of the pathways

that execute cellular processes. Approaches to do this combine systematic gene inhibition with functional tests that span a continuum—from single readout assays to complex assays that interrogate a broad spectrum of cellular processes. Whereas single readout assays identify pathways that impact a specific process (Mathey-Prevot and Perrimon, 2006), complex assays can be used to construct functional networks from collections of genes with diverse cellular roles. Two approaches have emerged for distilling complex phenotypes for phenotypic profiling: genetic interaction profiling and high-content screening. Although the methodologies are distinct, both strategies translate the consequences of inhibiting gene activity into phenotypic profiles that can be compared to generate a map of the functional relationships between genes (Boone et al., 2007; Collins et al., 2009; Conrad and Gerlich, 2010; Piano et al., 2002; Sönnichsen et al., 2005).

Genetic interaction profiling was pioneered in budding yeast, using a comprehensive deletion library of non-essential genes and collections of hypomorphic alleles of essential genes (Boone et al., 2007; Collins et al., 2009). Genetic interaction profiling captures the consequences of inhibiting a gene by measuring the effect on growth rate of pairwise inhibitions with each of the other genes in the collection. This analysis generates quantitative interaction profiles for each gene that can be clustered to reveal functionally significant relationships. A genome-scale genetic interaction map was recently constructed for *S. cerevisiae* (Costanzo et al., 2010), and maps have also been generated for subsets of gene implicated in specific processes—such as RNA processing, chromosome biology, proteasome function, and the secretory pathway (Breslow et al., 2008; Collins et al., 2007; Schuldiner et al., 2005; Wilmes et al., 2008).

In metazoans, genetic interaction profiling is difficult to implement because comprehensive libraries of deletion/hypomorphic strains do not exist and developing reproducible high-throughput methods to quantify fitness is a formidable barrier (Gunsalus, 2008). Consequently, high-content screening is the primary method for mapping functional gene networks in animal cells. In a high-content screen, light microscopy is used to assess phenotypes arising from gene inhibition by RNA-mediated interference, and the phenotype is captured by scoring a large parameter set (Conrad and Gerlich, 2010). The depth of the phenotypic profile is based on the biological complexity of the assay and the nature and accuracy of parameter scoring. To date, the requirement for high-resolution imaging has generally limited high-content screening to single cells or early embryos.

C. elegans is a prototype metazoan system for the functional mapping of essential genes (Piano et al., 2006). *C. elegans* has ~20,000 genes, of which ~2,500 are essential for embryo production or viability (WormBase release WS210; Harris et al., 2010). In a set of pioneering high-content screens, time-lapse Differential Interference Contrast (DIC) microscopy was used to film the early divisions of embryos following individual inhibitions of specific subsets of *C. elegans* genes (Gönczy et al., 2000; Piano et al., 2000; Zipperten et al., 2001). This was extended to a full-genome screen that generated high-content phenotypic profiles for ~500 essential genes (Sönnichsen et al., 2005). These profiles were combined with protein-protein interaction and expression profiling data to create a first-generation integrative map that linked 305 essential *C. elegans* genes in a “multiple support” network that grouped genes into modules involved in specific processes including spindle assembly, chromosome segregation, nuclear envelope dynamics, cortical dynamics, and centrosome function (Gunsalus et al., 2005). Despite the success of these studies, a large collection of essential genes could not be profiled because their inhibition results in sterility of the treated worm. Thus, the 554 genes in the “sterile” collection, which control fundamental cellular processes such as membrane trafficking, translation, proteasome function, and cortical remodeling, were largely absent from this analysis.

To fill this gap in the analysis of the *C. elegans* essential gene set, we profiled the 554 sterile genes by imaging syncytial gonad architecture at high-resolution following gene knockdown and scoring 94 phenotypic parameters. To generate a reference for evaluating computational methods for network construction, genes were manually partitioned into 102 phenotypic classes, predicting functions for 106 of the 116 uncharacterized genes in the collection. Using the manual classification as a benchmark, we developed a robust computational method for constructing gene networks from high-content profiles based on a network context-dependent measure that ranks the significance of functional links between genes. This method allowed us to integrate our data with that from the prior high-content embryo-filming dataset to generate a network representation of 818 essential *C. elegans* genes that can be viewed at multiple levels of functional resolution.

RESULTS

Phenotypic Profiling Based on The Morphology of a Complex Tissue: The *C. elegans* Gonad

Around 900 *C. elegans* genes are required for embryo production and/or for the early embryonic cell divisions (Sönnichsen et al., 2005); this collection includes the majority of genes essential for basic processes common to all cells. Because their inhibition leads to sterility, 554 of these genes could not be profiled by embryo filming. Of the 554 sterile genes, 166 were unnamed, indicating no prior characterization (Figure 1A). For each unnamed gene, we determined whether the predicted product is a member of a KOG (eukaryotic orthologous group; Tatusov et al., 2003) and used the Ensembl database to determine if it has orthologs across species. Of the 166 unnamed genes, 50 had characterized orthologs that predicted a function for the *C. elegans* protein (Unnamed-Group I in Table S2). The remaining 116 were either members of KOGs of unknown function, had no predicted orthologs, or had multiple *C. elegans* paralogs (Unnamed-Group II in Table S2); we refer to these 116 genes as “uncharacterized” (Figure 1A).

To profile the 554 sterile genes, we developed a high-content assay based on 3D two-color fluorescence confocal imaging of the gonad, a complex tissue in the adult *C. elegans* hermaphrodite (Figure 1B). The syncytial gonad contains ~1000 meiotic nuclei in cup-shaped compartments open to a common cytoplasmic core. Compartments mature into oocytes as they progress from the distal tip to the proximal region of the gonad adjacent to the spermatheca. Gonad maturation and maintenance involves a broad spectrum of basic cellular processes (Figure 1C), making this tissue an attractive substrate for high-content profiling. Gonad architecture was analyzed in a strain co-expressing fluorescent markers that target to the plasma membrane (GFP fusion that binds PI4,5P₂) and chromosomes (mCherry-histone H2B). Hermaphrodites were soaked in dsRNA against a target gene for 24 hr beginning at the late L4 larval stage, when the gonad has almost achieved its full complement of nuclei (Kimble and Crittenden, 2005). After 48 hr recovery, gonad architecture was assessed in triplicate by anesthetizing worms and imaging one gonad per worm.

Binary phenotypic profiles were generated by scoring the set of 3 image stacks per target gene for 94 possible defects. The movie set for each gene was inspected for each of the 94 defects (Figure 1D; for a complete list with examples see Table S1), assigning a “0” when the defect was absent and a “1” when the defect was present in at least 2 of the 3 movies. All image stacks were analyzed by the same pair of individuals, who viewed and scored them together; image stacks were indexed by RNA number, making their analysis blind to gene identity. In the 24 cases where the three movies were not consistent, the experiment was repeated (see Figure S1 available online, Extended Experimental Procedures, and Table S6 for details on screen design and scoring methods).

Generation of a Benchmark for Constructing Gene Networks Based on Gonad Architecture Phenotypes

Initial attempts using the raw parameter dataset for automated clustering broadly grouped genes, but failed to partition them

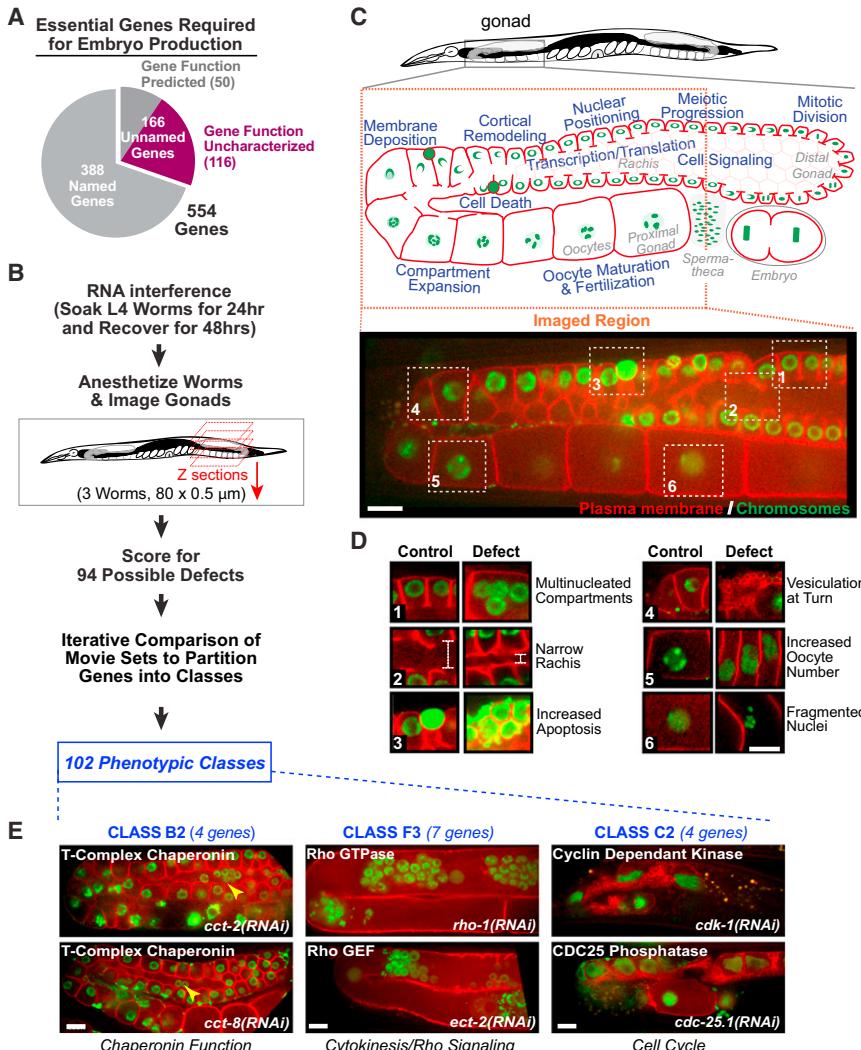


Figure 1. Phenotypic Profiling Using a High-Content Assay Based on the Architecture of a Complex Tissue

(A) Breakdown of the 554 genes in the sterile collection.

(B) Screen flow chart.

(C) The *C. elegans* gonad is a complex dynamic tissue whose architecture depends on a broad spectrum of interacting cellular processes (blue text on the schematic). Spinning disc confocal microscopy was used to collect an 80 plane two-color z-series of each gonad. The imaged region is indicated (red dashed box) and a sample central z-section from a control gonad is shown.

(D) Six sample defects are illustrated by pairing the numbered boxed regions from the control gonad in (C) with the corresponding regions from gonads with the indicated defects.

(E) Central plane images from two gene knock-downs in three phenotypic classes. Knock-downs in Class B2 (left column), which contains chaperonin complex subunits, led to rounded compartments and nuclei that fell out of their compartments (yellow arrowheads). Knock-downs in Class F3 (middle column), which contains genes implicated in Rho GTPase signaling, led to “tubulated” gonads with clustered nuclei. Knock-downs in Class C2 (right column), which contains cell cycle regulators, led to gonads with few compartments/oocytes. Bars represent 10 μm. See also Figure S1, Table S1, and Table S2.

at a resolution similar to what could be achieved through blinded manual classification by an experienced investigator. To develop a better computational method, we began by manually partitioning the genes into classes to generate a reference that we could use to evaluate computational methods for network construction. Manual partitioning placed the 554 sterile genes into 102 phenotypic classes (Table S2 contains a description, sample image, and gene list for each class). For organizational purposes, the 102 classes were grouped into 29 broad categories (labeled A-Z, AA, AB, and AC) that each contain classes sharing one or more prominent defects. Movies and class designations can be accessed via the Phenobank website (http://worm.mpi-cbg.de/phenobank_gonad).

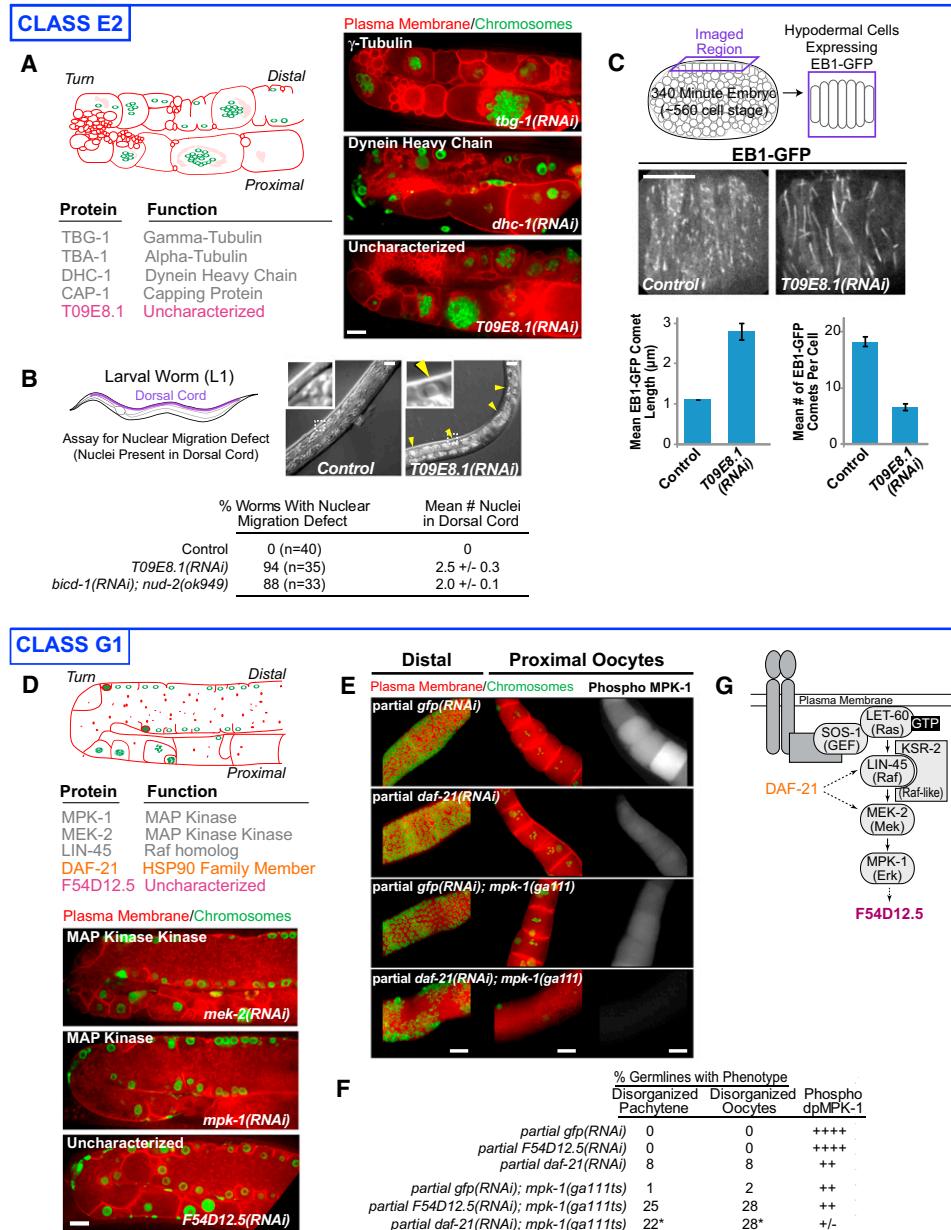
The manually defined phenotypic classes contained characterized genes with common annotated molecular functions (Figure 1E), indicating that similar gonad architecture phenotypes reflect similar molecular functions. Manual partitioning placed 106 of the 116 uncharacterized genes into phenotypic classes containing characterized genes, leading to predictions for their

functions. Prior to using the manually defined classes as a benchmark to optimize computational methods for network analysis, we validated the functional predictions arising from our classification by following up on 8 uncharacterized genes in 5 classes.

The first class (E2) predicted a role for T09E8.1 in the microtubule cytoskeleton

(Figure 2A). Unlike other class E2 genes, partial T09E8.1 inhibition did not lead to defects in the early embryonic cell divisions (not shown). In dividing cells, the microtubule cytoskeleton is organized by centrosomes, whereas the gonad is dominated by non-centrosomal microtubule arrays (Zhou et al., 2009). Thus, we hypothesized that T09E8.1 is specifically required for non-centrosomal microtubule array formation. Hypodermal cells, born late in embryogenesis, also have non-centrosomal microtubule arrays that function in nuclear positioning (Fridolfsson and Starr, 2010). Partial T09E8.1 inhibition in the hypodermis led to defects in nuclear positioning (Figure 2B), as well as a ~3-fold reduction in the number and an increase in the length of the GFP-EBP-1 “comets” that mark growing microtubule ends (Figure 2C). We conclude that T09E8.1, which we name *noca-1* (for non-centrosomal array), plays an important role in the organization of the noncentrosomal microtubule arrays in the gonad and hypodermis.

The second class (G1) predicted a role for F54D12.5 and DAF-21, an hsp90-family chaperone, in the MAPK signaling pathway

**Figure 2. Validation of Manual Classification I**

- (A) List of the genes in Class E2. Schematic and central plane images illustrate the class phenotype.
- (B) Nuclei are in the dorsal cords of *T09E8.1(RNAi)* (yellow arrowheads), but not control, L1 worms (insets 2.4×), reflecting a hypodermal cell nuclear migration defect. Table shows quantification of the nuclear migration defect. The effects of simultaneous inhibition of the dynein-regulatory proteins BICD-1 and NUD-2 quantified from the dataset in Fridolfsson et al. (2010) are shown for comparison.
- (C) The effect of *T09E8.1* knockdown on microtubule arrays in the hypodermis (schematic) was monitored by timelapse imaging of an EB1-GFP fusion (Fridolfsson and Starr, 2010). Bar graphs show the mean length (left) and number (right) of EB1-GFP comets. Error bars represent the SE.
- (D) Class G1 contains three characterized genes implicated in MAPK signaling (gray), 1 characterized gene not previously implicated in MAPK signaling (orange), and 1 uncharacterized gene (purple). Schematic and central plane images illustrate the class phenotype.
- (E) Partial RNAi of *daf-21*, or GFP as a control, was performed in the presence or absence of a weak *mpk-1* loss-of-function mutation (*ga111ts*). Gonads were fixed and stained for chromosomes (DAPI, green), the plasma membrane (SYN-4 and PTC-1, red), and activated MPK-1 (Phospho-MPK-1, right panels). All analysis was performed in the *rrf-1(pk1417)* background in which RNAi is effective in the gonad, but not in the surrounding somatic cells.
- (F) Quantification of phenotypes resulting from partial RNAi of *daf-21* or *F54D12.5* in the *mpk-1(ga111)* background (see Table S3). *Percentages do not include the 21% of germlines that showed a severe MPK-1 "null" phenotype.
- (G) Schematic places DAF-21 and F54D12.5 in the MAPK signaling pathway. Error bars are the SE. Bars, 10 μm. See also Figure S2, Table S2 and Table S3.

(Figure 2D). We tested these predictions using a genetic approach. Worms homozygous for the reduction-of-function allele *mpk-1(ga111)* exhibit normal gonad morphology, even though phosphorylated MPK-1 levels, a readout for pathway activity, are reduced compared to controls (Figures 2E and 2F; Lee et al., 2007). Partial knockdown of *daf-21* or F54D12.5, under conditions that did not result in a morphological phenotype in control worms, led to a strong MAPK knockdown phenotype in the *mpk-1(ga111)* background (Figures 2E and 2F; Figure S2, Table S3). RNAi of *daf-21* reduced phosphorylated MPK-1 levels in control and *let-60* gain-of-function worms (Figures 2E and 2F; Table S3), indicating that DAF-21 acts at the level of or downstream of LET-60/Ras in the MAPK pathway (Figure 2G). F54D12.5 RNAi did not reduce phosphorylated MPK-1 levels, and F54D12.5 contains 2 potential MPK-1 docking sites suggesting that it is an MPK-1 substrate (Figure 2G). We conclude that DAF-21 and F54D12.5 function at different points in the MAPK signaling pathway and name F54D12.5, *eom-1*, for enhancer of *mpk-1(ga111)*.

The third class (S2) predicted roles for three uncharacterized proteins in the anaphase-promoting complex/cyclosome (APC/C; Figure 3A). Immuno-affinity purification of one uncharacterized gene product, K10D2.4, from *C. elegans* extracts followed by mass spectrometry recovered seven APC components (Figure 3B). The product of another uncharacterized Class S2 gene, C09H10.7, was also recovered—indicating that both K10D2.4 and C09H10.7 are APC/C subunits. K10D2.4 was recently identified as a metazoan-specific component of the APC/C (Hubner et al., 2010; Hutchins et al., 2010; Kops et al., 2010) and the gene was named *emb-1/apc-16*. We name C09H10.7, *apc-17*.

The fourth class (F2) predicted a role for the BTB-domain containing protein C08C3.4 in cortical remodeling/cytokinesis (Figure 3C). A GFP fusion with C08C3.4 localized to the contractile ring at the tip of the cleavage furrow in dividing embryos (Figure 3D), and embryos partially depleted of C08C3.4 exhibited cytokinesis defects (Figure 3E). We conclude that C08C3.4 is required for cortical remodeling in the gonad and cytokinesis in embryos and name the gene *cyk-7*.

The fifth class (I2) phenotype included debris labeled with the plasma membrane probe, suggesting a role in membrane trafficking (Figure 3F). We tested for a trafficking function by imaging compartment boundaries in a strain co-expressing a mCherry labeled plasma membrane probe and a GFP fusion with SNB-1, a SNARE trafficked through the endomembrane system and delivered to the plasma membrane (Figure 3G). Compartment boundaries in control worms have SNB-1-GFP and the plasma membrane probe and are yellow. Trafficking defects prevent SNB-1-GFP from reaching the plasma membrane, leading to red compartment boundaries. 8 of the 15 Class I2 genes, including F27C8.6 and T01B7.6, exhibited defects in the SNB-1 assay (Figure 3G). We name F27C8.6 and T01B7.6 *trcs-1* and *trcs-2*, respectively, for (transport to the cell surface).

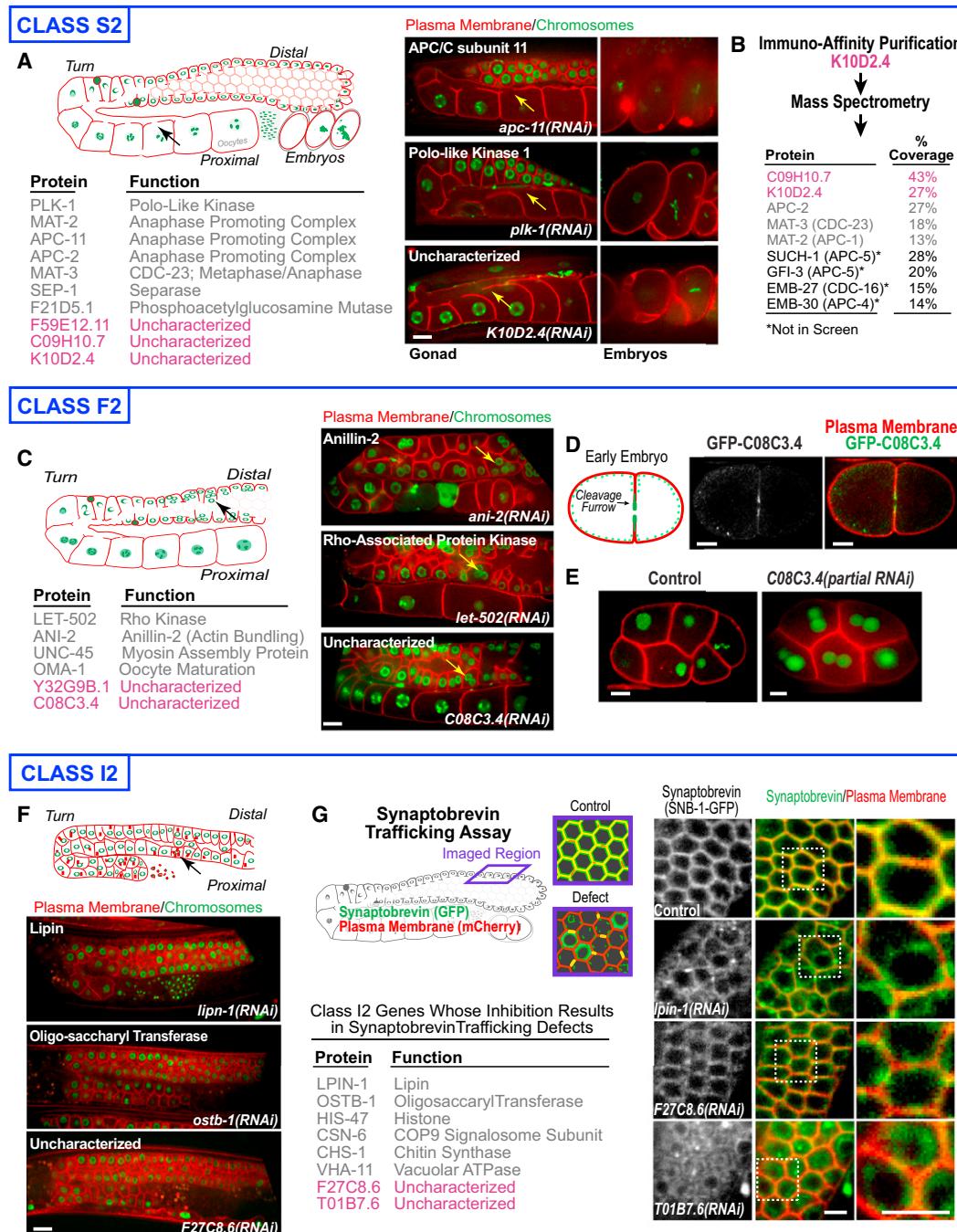
The follow up work on these five classes demonstrates that gonad architecture has sufficient resolution to functionally classify genes across a broad spectrum of essential cellular functions. It also validates the manual classification, establishing it as a benchmark for evaluating computational methods for network construction.

Development of a Network Context-Dependent Method to Evaluate the Significance of Functional Links Between Genes Based on High-Content Phenotypic Profiles

Our initial efforts using automated clustering were unable to partition genes at a resolution comparable to what could be achieved by an experienced investigator. To circumvent this limitation, we used the manually-defined classes as a tool to develop a robust computational method for constructing gene networks based on high-content parameter profiles. We first compared the phenotypic profiles by calculating the Pearson's Correlation Coefficient (PCC) for each pair of genes. The resulting network was visualized using N-Browse, an interactive Java-based tool (Kao and Gunsalus, 2008), to display connections between genes whose profiles were correlated with a PCC greater than or equal to a specified threshold (Figures 4A and 4B; dark blue lines connecting gray gene nodes). To assess the effectiveness of this approach, we circled gene clusters that corresponded to our manually-defined classes. This approach revealed that the optimal PCC threshold for viewing functionally relevant connections (red outlined boxes in Figure 4B) varied substantially between different network neighborhoods. This variability was due to the varying nature of the profiled phenotypes and the extent to which they are captured by the parameter set, the extent to which scored parameters are related versus independent, and the fact that profiles with more features often exhibit more variance. Viewing the entire network at a single PCC threshold is not possible because for some regions the threshold is too low and the view is cluttered with non-specific connections (Figure 4B, images above the red boxed images), and for other regions the threshold is too high, yielding an empty network in which many meaningful connections are absent (Figure 4B, images below the red boxed images).

To circumvent the limitations of PCC-based analysis, we developed a measure that ranks the significance of functional links between genes based on network context. For each pair of genes A and B, we assigned a Connection Specificity Index (CSI) by: (1) calculating the PCC for the connections between A and B and each of the other genes in the dataset, (2) counting the number of genes connected to A or B with $PCC \geq PCC_{AB} - 0.05$ (i.e., at a level comparable to or better than the correlation between A and B, correlations with a PCC up to 0.05 less than PCC_{AB} were considered similar—this offset was determined empirically); (3) dividing this number by the total number of genes in the screen (554); and (4) subtracting the result from 1.0 (Figure 4C). The CSI is equivalent to the fraction of genes in the dataset whose profiles are less similar to those of A and B than the profile of A is to the profile of B. For example, a CSI of 0.97 means that the similarity between A and B is highly specific: only ~3% of gene knockdowns have profiles with comparable or higher similarity to either A or B. Since the CSI scales uniformly with functional significance across the entire network, connections of a similar level of significance can simultaneously be displayed at a single CSI threshold (Figure 4D).

We evaluated the performance of CSI and PCC threshold networks by comparing the ability of an automated clustering algorithm (MINE, Module Identification in NEworks;

**Figure 3. Validation of Manual Classification II**

- (A) List of the genes in Class S2. Schematic and central plane images of gonads (left) and adjacent embryos (right) illustrate the class phenotype.
- (B) List of relevant proteins identified by mass spectrometry in an immuno-affinity purification of K10D2.4 from worm extracts, along with percent coverage. Proteins encoded by the uncharacterized Class S2 genes (pink) are listed, along with APC components in Class S2 (gray) and additional APC components not in the screen (black).
- (C) List of the genes in Class F2. Schematic and central plane images illustrate the class phenotype.
- (D) Embryo co-expressing GFP-C08C3.4 (green) and an mCherry tagged plasma membrane probe (red) during the first cell division.
- (E) Central plane images of control and C08C3.4(RNAi) embryos expressing mCherry-histone H2B (green) and a GFP plasma membrane probe (red). Multiple nuclei in each cell in the C08C3.4(RNAi) embryo are due to cytokinesis failure.
- (F) Schematic and central plane images illustrate the Class I2 phenotype, which includes punctate debris containing the plasma membrane probe (arrow in schematic).
- (G) Schematic of the trafficking assay. List of Class I2 genes with defects in the SNB-1-GFP trafficking assay. Images of the assay for 1 characterized and 2 uncharacterized Class I2 genes. Bars represent 10 μm in (A)–(F) and 5 μm in (G). See also Table S2.

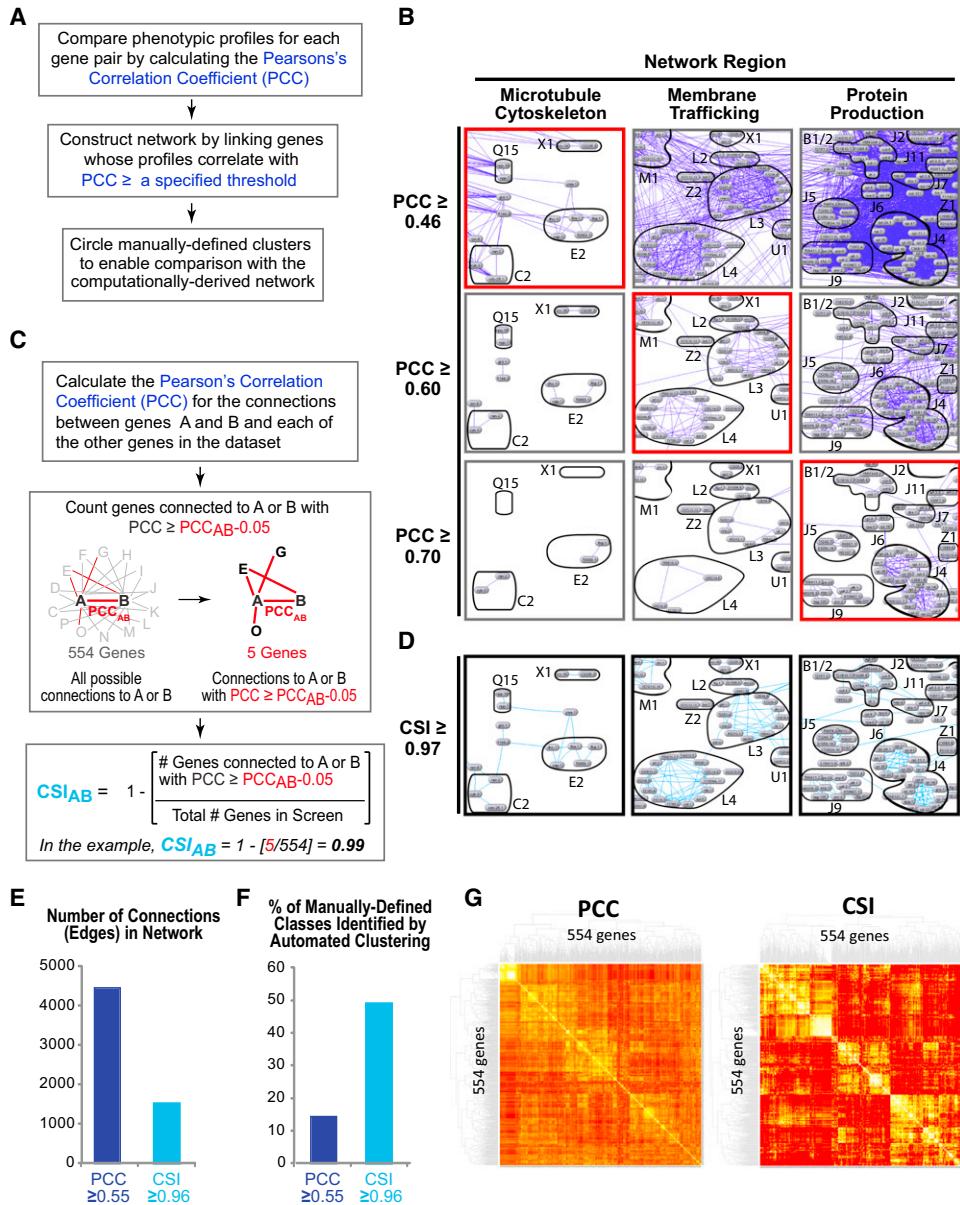


Figure 4. Constructing Gene Networks Using the Connection Specificity Index Instead of the PCC Reduces Connection Noise and Allows Connections of Similar Functional Significance to Be Viewed Across the Entire Network at a Uniform Threshold

(A) Flowchart of the steps used to construct the gene networks in (B).

(B) Gene networks were constructed by displaying connections (dark blue lines) between genes whose knockdown profiles were correlated with a $PCC \geq$ three specified thresholds (0.46, top; 0.60, middle; 0.70, bottom). Each column shows a network region, labeled based on the primary function of the genes in that region. To compare the computational network to the manually-defined phenotypic classes, gene groups from manually-assigned classes were circled and labeled. The optimal PCC threshold at which significant connections were displayed and non-specific connections were filtered out (red boxes) was different for different network regions.

(C) Method used to calculate the CSI.

(D) Gene network showing the same regions in (B) constructed by displaying connections (light blue lines) with a $CSI \geq 0.97$.

(E) Bar graph showing the number of connections in gene networks constructed using PCC or CSI thresholds of 0.55 or 0.96, respectively. These thresholds were chosen because they are the highest thresholds that retain most of the genes in the network; genes drop out of the network when they no longer make any connections with a PCC/CSI that exceeds the specified threshold.

(F) Bar graph showing the percent of the 49 manually-defined phenotypic classes containing 4 or more genes identified by an automated clustering algorithm (MINE) in networks constructed using PCC or CSI thresholds of 0.55 or 0.96, respectively.

(G) Heatmap dendograms of the sterile gene set constructed based on the PCC or the CSI. See also Figure S3 and Table S4.

Rhrissorakrai and Gunsalus, in press) to identify gene clusters corresponding to our manually-defined classes. This analysis revealed that over the useful range of the two parameters, networks generated using CSI thresholds have ~3-fold fewer connections than comparable PCC networks; this noise reduction translates into a substantial improvement in the ability of an automated algorithm to identify functionally relevant gene clusters (Figures 4E and 4F; Figure S3). Approximately half of the manually-defined classes containing 4 or more genes could be identified by automated clustering in a network constructed using the single CSI threshold of 0.96, whereas only 14% could be identified in a comparable PCC network (Figure 4F). When MINE was allowed to search networks spanning a range of CSI thresholds (0.90 to 0.99), clusters corresponding to ~90% of the manually-defined classes could be identified, compared to 65% for networks spanning a range of PCC thresholds (0.5-1). Heat map dendograms also revealed more sharply defined clusters when using the CSI, indicating that the CSI increases network clarity (Figure 4G; Table S4).

We conclude that constructing networks using CSI thresholds circumvents the variability in the significance of a strict correlation measure in different network regions that arises due to the complexity of high-content screening parameters. Using the CSI instead of the PCC reduces connection noise and allows connections of a similar functional significance to be viewed across the entire network at a single threshold.

Varying the CSI Threshold Reveals Functional Modularity in the Gene Network at Different Levels of Resolution

The CSI-based network representation allows exploration of functional modularity at different levels of resolution. This point is illustrated by the region of the gene network involved in protein production (Figure 5A): a relatively low threshold CSI of 0.90 connects the entire set of genes involved in protein translation, mRNA splicing, and protein folding in a dense meshwork. An intermediate CSI threshold of 0.93 results in a sparser network of more specific connections that defines smaller gene groups. At a high CSI threshold of 0.97, the chaperonins, tRNA synthetases, translation initiation/elongation factors, small ribosome subunits, large ribosome subunits, and splicing factors are resolved into separate clusters. Thus, dialing the CSI up or down reveals functional relationships at different levels of resolution. The automated clusters identified by MINE at three different CSI thresholds are provided in the supplement (Table S5; note that genes can be in multiple clusters).

We assessed the resolution limit of the gonad architecture assay using the region of the network representing genes involved in protein degradation (Figure 5B, Figure S4). Above the very high CSI threshold of 0.99, the connections that remained linked genes within specific proteasome subcomplexes (the Lid, the core β -ring, the core α -ring or the 19S ATPase base; Figure 5B). We conclude that phenotypic profiling based on complex tissue architecture, coupled with automated construction of CSI-based networks from parameter profiles, is capable of correctly assigning very fine distinctions in protein function.

CSI-Enabled Integration of High-Content Data Sets Generates a Global View of the *C. elegans* Essential Gene Network

Together, the gonad architecture (554 genes) and time-lapse embryo filming (661 genes) screens provide high-content profiles for 885 essential genes (330 were profiled in both screens). The CSI is ideally suited for integrating these datasets because it filters out low specificity phenotypic links, leading to a network that combines only the significant relationships identified by the two screens. We binarized the phenotypic signatures in the embryo timelapse screen, which were based on scoring for 45 possible defects (Sönnichsen et al., 2005), calculated a CSI for each gene pair, and simultaneously displayed both data sets in N-Browse to create an integrated network. The merged network, or network regions centered on genes of interest, can be viewed at any CSI threshold in N-Browse (instructions and a demo video describing how to use N-Browse can be accessed at http://worm.mpi-cbg.de/phenobank_gonad/nbrowse). At a CSI threshold of 0.96, the integrated network has 3382 high significance connections linking 818 genes (Figures 6A–6C). The combined view provides functional information for ~90% of the genes required for embryo production or the first two embryonic divisions. Despite the presence of 330 genes in common, the relationships identified by the two datasets were highly orthogonal—only 23 of the total 3382 phenotypic links are shared (Figure 6A). This lack of overlap is explained by the majority of the 330 common genes failing to make high-significance connections in the embryo time-lapse screen due to sterility onset (Figure S5). Overall, the CSI-based network integrates information acquired in the biologically distinct contexts of the gonad and embryo to provide a multi-layered view of the function of 818 genes required for essential cellular processes in a multicellular organism.

DISCUSSION

High-content screening is the primary method for mapping functional gene networks in animal cells (Collins et al., 2009; Conrad and Gerlich, 2010; Gunsalus, 2008). A number of high-content screens have classified genes based on cell morphology following gene knockdown (examples include Bakal et al., 2007; Echard et al., 2004; Eggert et al., 2004; Goshima et al., 2007; Liu et al., 2009; Neumann et al., 2010; Piano et al., 2002; Sönnichsen et al., 2005). While successful in identifying genes that contribute to cell division and morphology, the ability of these screens to assess protein function across a breadth of cellular processes falls far short of what has been attained using genetic interaction profiling in fungi. Screen resolution—the ability to discriminate between different biological functions—depends on the information content of the phenotypic assay; therefore, one approach to enhance resolution has been time-lapse imaging (Conrad and Gerlich, 2010). Here, we explored an alternative approach—analyzing the effects of gene knockdowns on complex tissue architecture at a single time point in a multicellular organism. We monitored *C. elegans* gonad architecture following gene knockdown by scoring 94 phenotypic features. Our results demonstrate that the biological complexity of the gonad translates into phenotypes that enable profiling across a wide spectrum of essential cellular processes at a resolution approaching that of genetic interaction profiling in yeasts.

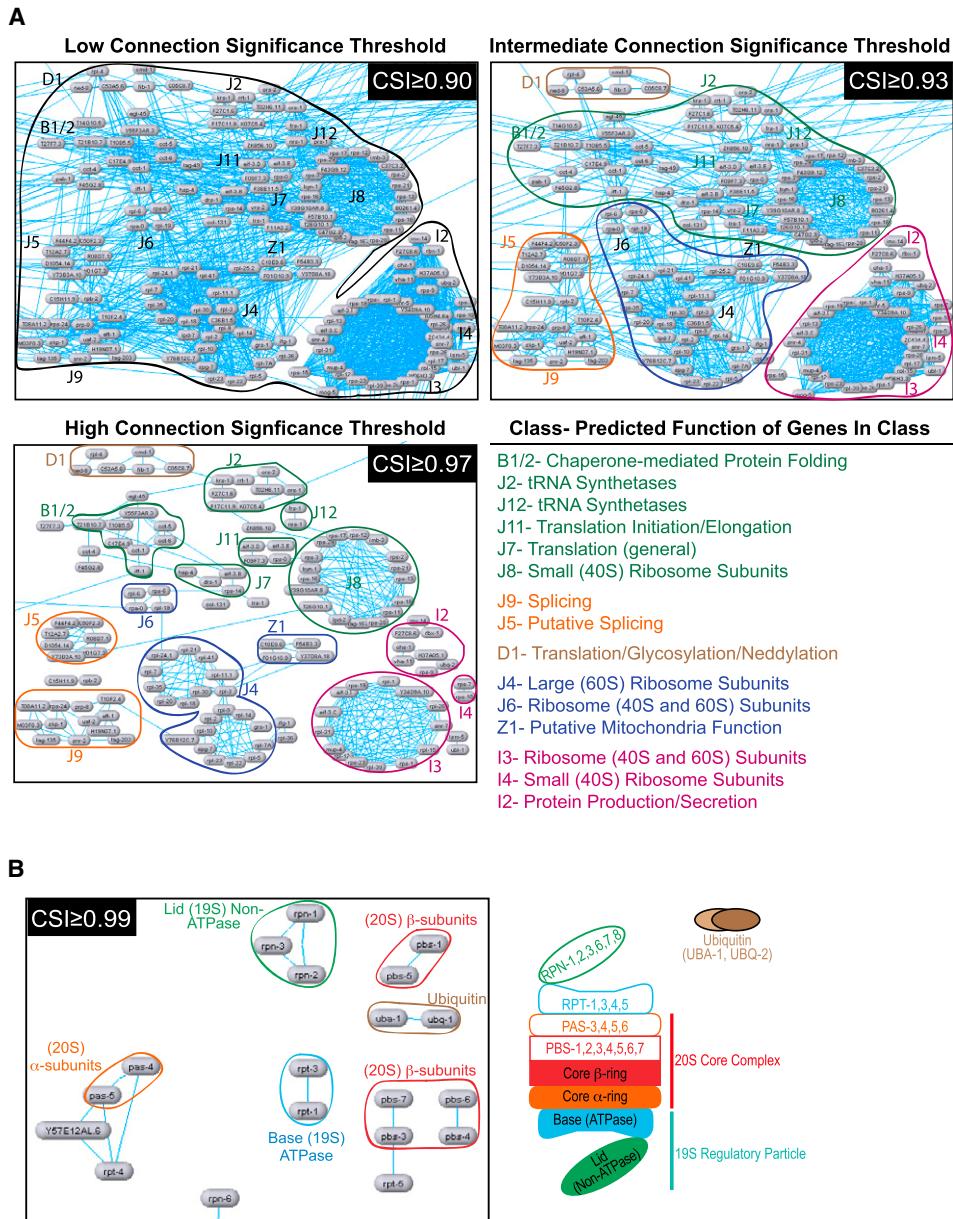


Figure 5. Varying the CSI Threshold Reveals Functional Modularity in the Gene Network at Different Levels of Resolution

(A) The region of the gene network involved in protein production is shown using three different CSI thresholds to filter displayed connections. The gene clusters apparent at each threshold are circled and gene groups from manually-defined classes are labeled.

(B) The region of the network involved in protein degradation is shown using the very high CSI threshold of 0.99. The connections that remain link components within specific proteasome subcomplexes (illustrated schematically on the right).

See also Figure S4 and Table S5.

The Sterile Gene Collection: A Prominent Gap in the Functional Genomic Analysis of the *C. elegans* Essential Gene Set

Sequencing of the *C. elegans* genome and the discovery of RNAi catalyzed efforts to systematically catalog the functions of its predicted genes (Piano et al., 2006). *C. elegans* has ~900 essential genes required for embryo production and/or events during the first two embryonic divisions. A subset of these genes was

previously characterized by time-lapse imaging of early embryos (Fraser et al., 2000; Gönczy et al., 2000; Piano et al., 2002; Sönnichsen et al., 2005). However, sterility onset following RNAi was a major complication that precluded in-depth characterization of a large number of genes. Our efforts focused on these 554 sterile genes, which we profiled by imaging gonad architecture following RNAi knockdowns and generating a parameterized description of the resulting phenotypes.

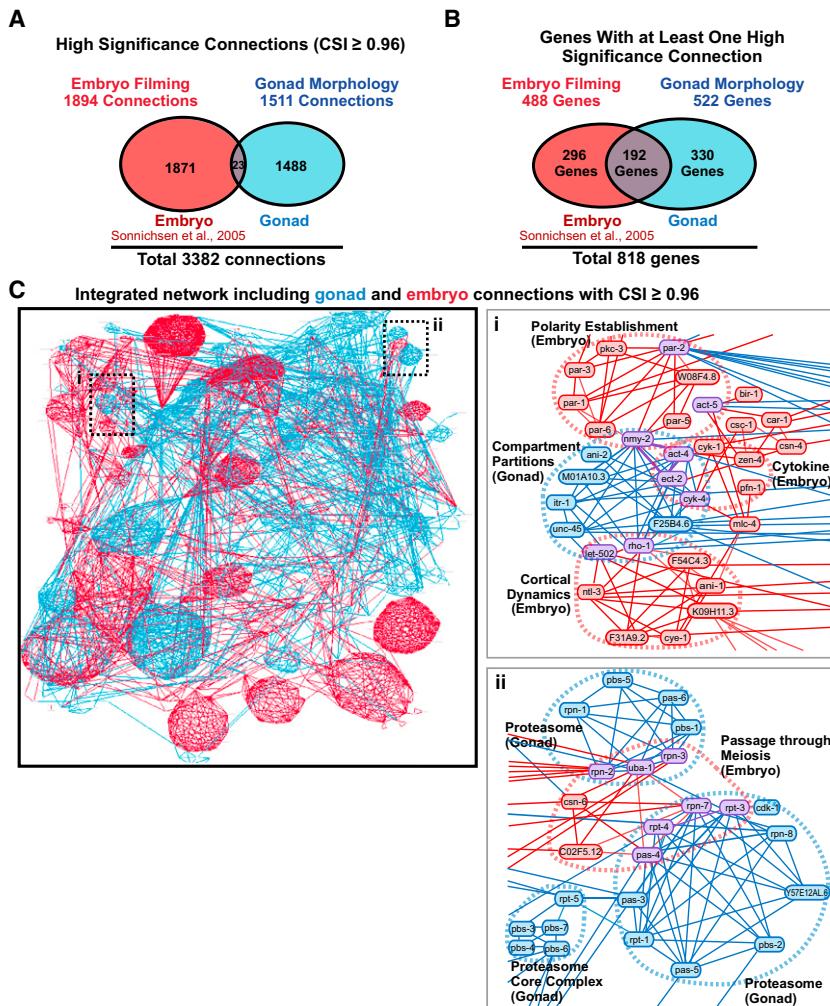


Figure 6. The CSI Enables Integration of High-Content Data Sets to Generate a Global View of the Essential *C. elegans* Gene Network

(A) Venn diagram showing the high significance connections identified by the embryo-filming (Sönnichsen et al., 2005) and gonad architecture data.

(B) Venn diagram showing the genes that make at least one high significance connection identified by each dataset.

(C) Bird's eye view of the integrated network combining high-significance connections based on the gonad (blue) and embryo (red) data. Insets (i-ii) highlight regions where the gonad and embryo data intersect (connections identified by both datasets are purple).

See also Figure S5.

High-Resolution Phenotypic Profiling Based on the Morphology of a Complex Tissue

While the original motivation for our screen was to profile the sterile gene set, a fortuitous byproduct was the realization that a complex tissue, whose structure depends on the dynamic coordination of a broad spectrum of cellular processes, is an ideal substrate for phenotypic profiling. A major finding of our work is that profiles based on complex tissue architecture acquired at a single time point following gene knockdown have greater information depth than profiles derived from timelapse imaging of the first two divisions of the early embryo (a single cell). For example, as shown in Figure 5, the gonad data partitions the set of genes involved in protein

production into 15 different classes (chaperones, ribosome components, tRNA synthetases, proteins involved in protein folding, splicing components etc.) whereas none of these distinctions could be made based on embryo filming data. This high level of resolution generally holds throughout the dataset.

We note that the means of inferring function in a single point morphology assay is distinct from that used in the embryo-filming screen. In the embryo screen, gene function was inferred directly from phenotype. For example, defects in spindle assembly, chromosome segregation, or polarity establishment led to assignment to classes implicated in the corresponding processes (Sönnichsen et al., 2005). By contrast, the link between catalogued phenotypic features in the gonad architecture assay and gene function is not direct—we do not know why knockdown of genes with specific functions lead to a specific spectrum of phenotypic features. Instead, protein function is inferred by comparing the profile of phenotypic features to those of other genes in the dataset, a process conceptually analogous to how function is inferred from genetic interaction profiles in budding yeast.

Our analysis generated functional predictions for 106 of the 116 uncharacterized sterile genes. These predictions span a variety of cellular processes including membrane trafficking, glycosylation, fatty acid synthesis, cortical dynamics, chromosome structure/segregation, RNA splicing, mitochondrial function, microtubule cytoskeleton, proteasome function, translation, MAPK signaling, transcription, RNA binding, and the metaphase-anaphase transition. We validated these predictions for 8 genes that play important roles in the microtubule cytoskeleton, MAP kinase signaling, the metaphase-anaphase transition, cortical remodeling/cytokinesis, and membrane trafficking. Although the gonad was used as a “test tube” to functionally profile these genes, subsequent work in early and later stage embryos indicates that the majority of the essential genes in this collection are broadly important. Of the 554 sterile genes, ~60% have predicted homologs in higher organisms, and more than 50% of the 106 uncharacterized genes have a predicted homolog or conserved domain. Thus, in addition to filling a large gap in the analysis of the *C. elegans* essential gene set, our findings provide a starting point to address the functions of related genes in humans.

The presence of proteasome components in our dataset offered an opportunity to compare the resolution of our approach to genetic interaction profiling in yeast. The proteasome is a 2.5-MDa macro-molecular machine that contains over 30 different subunits organized into a 20S core complex, composed of α and β rings, and a 19S regulatory particle, containing base and lid sub-complexes (Bochler et al., 1999). In yeast, distinct genetic interaction profiles were obtained for the subunits of the four proteasome sub-complexes using a sensitive competitive growth assay (Breslow et al., 2008). At a CSI ≥ 0.99 (a high-stringency filter), the connections that remained were between subunits of specific proteasome subcomplexes (α ring, β ring, lid and base; Figure 5B). This observation indicates phenotypic profiles based on gonad morphology provide resolution sufficient to group subunits of specific proteasome subcomplexes.

At the core of our dataset are profiles composed of parameters visually scored by experienced investigators rather than acquired through automated image analysis. Given the complexity of the substrate—a 3 dimensional tissue in a living organism that can be variably positioned in the worm—and the large spectrum of knockdown phenotypes that can entirely change the properties of the structure (gonad size, shape, position, compartment and nuclei number and morphology) automated parameter scoring would have been exceedingly difficult. However, the exact properties that make automated analysis difficult—the varied and dramatic effects that gene knockdowns can have on gonad architecture—are also the properties that give the assay its profiling power. Although manual parameter scoring could introduce some bias, this was minimized by performing the analysis blinded to gene identity and the fact that the individual parameters were scored by investigators who were largely oblivious to the larger patterns that would ultimately emerge.

A Robust Computational Method for Constructing Gene Networks From High-Content Screening Data

By evaluating computational methods for network construction using a validated manual classification of a rich phenotypic dataset, we were able to devise a robust computational method for constructing gene networks from high-content phenotypic profiles. This method overcomes two challenges encountered in network analysis of high-content datasets. The first challenge is that the level of profile correlation that is significant varies in different network regions due to the varying nature of the profiled phenotypes and the extent to which they are captured by the parameter set. The second challenge is that commonly encountered (and thus less informative) phenotypes can generate a “hairball” of connections that obscures meaningful functional links. At the center of the method we developed to overcome these challenges is a simple metric—the CSI, which is a network context-dependent measure that ranks the significance of functional links between genes. Compared to the Pearson’s Correlation Coefficient, constructing networks based on the CSI reduces non-specific connection noise, improves network clarity, and allows connections of a similar functional significance to be simultaneously viewed across the entire network at a single threshold.

Ranking connection significance allows exploration of the gene network at different levels of functional resolution and integration of high-content screening data from different sources.

We demonstrate the usefulness of the CSI by using it to integrate the high-content data from our gonad architecture screen with that from the prior embryo-filming screen, to generate an integrated network that provides a multi-layered view of 818 genes in the *C. elegans* essential gene set. The phenotypic profiles in these datasets are composed of parameters that were visually scored rather than measured through automated image analysis. However, constructing networks based on phenotypic profiles faces the same challenges, regardless of whether parameters are scored through manual or automated means. Consequently, we anticipate that the CSI-based method described here will be of equal utility in analyzing and integrating datasets composed of parameters acquired through automated analysis.

EXPERIMENTAL PROCEDURES

C. elegans Strains

Strains are listed in Table S7. The strains OD95, UD299, DP38, OD70, MSN142, NL2098, and BS3623 were previously described (Arur et al., 2009; Essex et al., 2009; Fridolfsson and Star, 2010; Kachur et al., 2008; Maduro and Pilgrim, 1995; Shi et al., 2010; Sijen et al., 2001). OD447 was generated by using a PDS-1000/He Biolistic Particle Delivery System (Bio-Rad Laboratories; Praitis et al., 2001) to bombard a construct containing the C08C3.4 genomic locus cloned into the Spel site of plC26 (Cheeseman et al., 2004) into DP38. OD449 was generated by mating OD447 with OD70.

RNA Production

Templates for dsRNA production were generated by using primers with tails containing the T3 and T7 promoters to amplify to amplify a 500–1000 bp region of the corresponding gene from genomic DNA. When possible, the oligo pairs used by Sönnichsen et al. (2005) were chosen. New oligos were designed for genes not in the Sönnichsen screen and when the Sönnichsen oligo pairs amplified introns or regions smaller or larger than 500–1000 bp (oligos are listed in Table S2). PCR reactions were cleaned (PerfectPrep 96 kit; Eppendorf) and used as templates for T3 and T7 transcription reactions (Megascript T3 and T7 kits, Ambion). Transcription reactions were mixed, cleaned (MegaClear96 kit, Ambion), and annealed by adding 3x Soaking buffer (32.7 mM Na₂HPO₄, 16.5 mM KH₂PO₄, 6.3 mM NaCl, 14.1 mM NH₄Cl) to a final concentration of 1X and incubating the reactions at 68°C for 10 min followed by 37°C for 30 min.

RNA Interference

Larval (L4 stage) worms from the strain OD95 were rinsed with M9 and soaked in 5 μ l of dsRNA (supplemented with 0.5 μ l of a 50/50 mixture of 63 mM spermidine and 1.1% gelatin) for 24 hr at 20°C inside a humid chamber. Worms were transferred to NGM plates seeded with OP50 *E. coli* and allowed to recover at 20°C for 48 hr prior to imaging.

Live Imaging

Six worms were anesthetized in a fresh mixture of 1 mg/ml Tricane (ethyl 3-aminobenzoate methanesulfonate salt) and 0.1 mg/ml of tetramisole hydrochloride (TMHC) dissolved in M9 for 15–30 min before transferring them to an agarose pad under a coverslip for imaging. Gonads were imaged by collecting an 80 \times 0.5 μ m z-series including DIC, GFP and RFP images for every z-plane. Microscopy was performed with a spinning disk confocal mounted on a Nikon TE2000-E inverted microscope equipped with a 60x 1.4 NA PlanApoChromat lens, a krypton-argon 2.5 W water-cooled laser (Spectra-Physics, Mountain View, CA) and a Hamamatsu Orca ER CCD camera. Acquisition parameters, shutters, and focus were controlled by MetaMorph software (Molecular Devices, Downington, PA).

Sterility Assessment

At the time when the other worms in the cohort were anesthetized and imaged, 3 worms were moved to a fresh plate. After 24 hr at 20°C, the number of

embryos and L1 progeny on the plate was counted. A gene was “sterile” if no embryos or L1 larvae were observed, “partially sterile” if < 50 embryos/L1 worms were present, and “wild-type” if > 50 embryos/L1 worms were present. After an additional 24 hr, we assessed whether the knockdown resulted in embryonic lethality. If > 50 L1-L2 larvae were present, they were considered “viable,” if < 50 L1-L2 larvae were present and there were > 5 dead embryos, it was considered “embryonic lethal.”

Antibody Production

Antibodies against K10D2.4 were generated by using the oligos cgcgcggatccgcgttgcattc and ggcgcgaaattctcaaggatttcaggcatatt to amplify a region encoding amino acids 2-81 from a cDNA library. The product was digested with BamHI/EcoRI and cloned into pGEX6P-1 digested with the same enzymes (GE Healthcare Life Sciences). The purified GST fusion protein was outsourced for injection into rabbits (Covance). Antibodies were affinity purified by binding to columns of the same antigen after removal of the GST tag as described previously (Desai et al., 2003).

Immunoprecipitation and Mass Spectrometry

K10D2.4 was immunoprecipitated from extracts prepared from frozen worm pellets (Desai et al., 2003) and mass spectrometry was performed (Cheeseman et al., 2004) as described, using the most recent version of the predicted *C. elegans* proteins (Wormprep111).

Hypodermal Cell Experiments, MAP Kinase Experiments, and Membrane Trafficking Assay

For Figure 2C, embryos produced by adult hermaphrodites from the strain UD299 were imaged as described (Fridolfson and Starr, 2010), 20 hr after injection of dsRNA against T09E8.1 into their gonad. Nuclear migration was scored in L1 larvae using DIC optics as described (Starr et al., 2001). The attenuated RNAi, gonad dissection, immunofluorescence, and gonad analysis in Figure 2E and F (Arur et al., 2009) and the GFP-SNB-1 trafficking assay (Shi et al., 2010) were performed as described.

Phenobank and N-Browse

Movies can be accessed via the Phenobank website (http://worm.mpi-cbg.de/phenobank_gonad) or through RNADB (<http://www.rnadb.org>). Written instructions and a demo video explaining how to search Phenobank can be found at http://worm.mpi-cbg.de/phenobank_gonad/project. Written instructions and a demo video describing how to visualize the gene network in N-Browse2 can be found at http://worm.mpi-cbg.de/phenobank_gonad/nbrowse. As described in the instructions, gonad and embryo data can be viewed in N-Browse2 by entering the URL <http://gnetbrowse.org/gonad.jnlp> into your browser and providing the password “phenotypes” at the prompt. Additional information about N-Browse, including system requirements and tutorials, can be found at <http://www.gnetbrowse.org>.

Automated clustering using MINE

MINE (Module Identification in Networks) is a graph clustering algorithm that shows comparable or better performance compared with similar algorithms in identifying functional modules in dense interaction networks (K. Rhrissorakrai and K.C. Gunsalus, *in press*). MINE, which is similar to MCODE (Bader and Hogue, 2003) but uses a modified weighting scheme and takes into account network modularity, is available as a Cytoscape plug-in from <http://www.cytoscape.org> or as a Perl package (upon request from the authors).

SUPPLEMENTAL INFORMATION

Supplemental Information includes Extended Experimental Procedures, five figures, and seven tables and can be found with this article online at doi:10.1016/j.cell.2011.03.037.

ACKNOWLEDGMENTS

This work was supported by grants from the American Cancer Society (PF-06-254-01-CCG, to RG), the Helen Hay Whitney Foundation (to A.A.), and the NIH

(R01 GM074207 to K.O., R01 HD046236 to K.C.G. and FP; R01 GM085503 to K.C.G.; R01 GM085150 to TS; R01 GM088151 to A.A.; R01 GM073874 to D.A.S.) and by funding from the Ludwig Institute for Cancer Research to KO and AD. We thank Kahn Rhrissorakrai for the MINE algorithm, Michael Volkmer for movie formatting, and the *Caenorhabditis* Genetics Center, funded by the NIH National Center for Research Resources, for strains. *Author Contributions:* Conceived and designed the experiments: KO, AD, RG, AA. Performed the experiments: RG, AA, SN, SA, TS, SW, HF, DS, JM. Scored the data: AA, RG, KO. Manually partitioned the genes into classes: RG, KO. Re-worked Phenobank to include the gonad morphology data: SS, RG, KO, AH. Developed N-Browse: HK, MS, FP, KG. Developed CSI: KO, RG, HK, MS, FP, KG. Conceptualized and performed network analysis: HK, FP, KG. Contributed reagents/materials/analysis tools: KO, RG, AA, SN, KL, SA, TS, AD, HK, MS, KG, FP, HF, DS.

Received: June 5, 2010

Revised: December 21, 2010

Accepted: March 24, 2011

Published: April 28, 2011

REFERENCES

- Arur, S., Ohmachi, M., Nayak, S., Hayes, M., Miranda, A., Hay, A., Golden, A., and Schedl, T. (2009). Multiple ERK substrates execute single biological processes in *Caenorhabditis elegans* germ-line development. Proc. Natl. Acad. Sci. USA 106, 4776–4781.
- Bader, G.D., and Hogue, C.W. (2003). An automated method for finding molecular complexes in large protein interaction networks. BMC Bioinformatics 4, 2.
- Bakal, C., Aach, J., Church, G., and Perrimon, N. (2007). Quantitative morphological signatures define local signaling networks regulating cell morphology. Science 316, 1753–1756.
- Bochtler, M., Ditzel, L., Groll, M., Hartmann, C., and Huber, R. (1999). The proteasome. Annu. Rev. Biophys. Biomol. Struct. 28, 295–317.
- Boone, C., Bussey, H., and Andrews, B.J. (2007). Exploring genetic interactions and networks with yeast. Nat. Rev. Genet. 8, 437–449.
- Boxem, M., Maliga, Z., Klitgord, N., Li, N., Lemmens, I., Mana, M., de Lichtenfelde, L., Mul, J.D., van de Peut, D., Devos, M., et al. (2008). A protein domain-based interactome network for *C. elegans* early embryogenesis. Cell 134, 534–545.
- Breslow, D.K., Cameron, D.M., Collins, S.R., Schuldiner, M., Stewart-Ornstein, J., Newman, H.W., Braun, S., Madhani, H.D., Krogan, N.J., and Weissman, J.S. (2008). A comprehensive strategy enabling high-resolution functional analysis of the yeast genome. Nat. Methods 5, 711–718.
- Cheeseman, I.M., Niessen, S., Anderson, S., Hyndman, F., Yates, J.R., 3rd, Oegema, K., and Desai, A. (2004). A conserved protein network controls assembly of the outer kinetochore and its ability to sustain tension. Genes Dev. 18, 2255–2268.
- Collins, S.R., Miller, K.M., Maas, N.L., Roguev, A., Fillingham, J., Chu, C.S., Schuldiner, M., Gebbia, M., Recht, J., Shales, M., et al. (2007). Functional dissection of protein complexes involved in yeast chromosome biology using a genetic interaction map. Nature 446, 806–810.
- Collins, S.R., Weissman, J.S., and Krogan, N.J. (2009). From information to knowledge: new technologies for defining gene function. Nat. Methods 6, 721–723.
- Conrad, C., and Gerlich, D.W. (2010). Automated microscopy for high-content RNAi screening. J. Cell Biol. 188, 453–461.
- Costanzo, M., Baryshnikova, A., Bellay, J., Kim, Y., Spear, E.D., Sevier, C.S., Ding, H., Koh, J.L., Toufighi, K., Mostafavi, S., et al. (2010). The genetic landscape of a cell. Science 327, 425–431.
- Desai, A., Rybina, S., Muller-Reichert, T., Shevchenko, A., Hyman, A., and Oegema, K. (2003). KNL-1 directs assembly of the microtubule-binding interface of the kinetochore in *C. elegans*. Genes Dev. 17, 2421–2435.
- Echard, A., Hickson, G.R., Foley, E., and O’Farrell, P.H. (2004). Terminal cytokinesis events uncovered after an RNAi screen. Curr. Biol. 14, 1685–1693.

- Eggert, U.S., Kiger, A.A., Richter, C., Perlman, Z.E., Perrimon, N., Mitchison, T.J., and Field, C.M. (2004). Parallel chemical genetic and genome-wide RNAi screens identify cytokinesis inhibitors and targets. *PLoS Biol.* 2, e379.
- Essex, A., Dammermann, A., Lewellyn, L., Oegema, K., and Desai, A. (2009). Systematic analysis in *Caenorhabditis elegans* reveals that the spindle checkpoint is composed of two largely independent branches. *Mol. Biol. Cell* 20, 1252–1267.
- Fraser, A.G., Kamath, R.S., Zipperlen, P., Martinez-Campos, M., Sohrmann, M., and Ahringer, J. (2000). Functional genomic analysis of *C. elegans* chromosome I by systematic RNA interference. *Nature* 408, 325–330.
- Gönczy, P., Echeverri, C., Oegema, K., Coulson, A., Jones, S.J., Copley, R.R., Duperon, J., Oegema, J., Brehm, M., Cassin, E., et al. (2000). Functional genomic analysis of cell division in *C. elegans* using RNAi of genes on chromosome III. *Nature* 408, 331–336.
- Fridolfsson, H.N., and Starr, D.A. (2010). Kinesin-1 and dynein at the nuclear envelope mediate the bidirectional migrations of nuclei. *J. Cell Biol.* 191, 115–128.
- Goshima, G., Wollman, R., Goodwin, S.S., Zhang, N., Scholey, J.M., Vale, R.D., and Stuurman, N. (2007). Genes required for mitotic spindle assembly in *Drosophila* S2 cells. *Science* 316, 417–421.
- Gunsalus, K.C. (2008). A *Caenorhabditis elegans* genetic-interaction map wiggles into view. *J. Biol.* 7, 8.
- Gunsalus, K.C., Ge, H., Schetter, A.J., Goldberg, D.S., Han, J.D., Hao, T., Berriz, G.F., Bertin, N., Huang, J., Chuang, L.S., et al. (2005). Predictive models of molecular machines involved in *Caenorhabditis elegans* early embryogenesis. *Nature* 436, 861–865.
- Harris, T.W., Antoshechkin, I., Bieri, T., Blasiar, D., Chan, J., Chen, W.J., De La Cruz, N., Davis, P., Duesbury, M., Fang, R., et al. (2010). WormBase: a comprehensive resource for nematode research. *Nucleic Acids Res.* 38, D463–D467.
- Hubner, N.C., Bird, A.W., Cox, J., Splettstoesser, B., Bandilla, P., Poser, I., Hyman, A., and Mann, M. (2010). Quantitative proteomics combined with BAC TransgeneOmics reveals *in vivo* protein interactions. *J. Cell Biol.* 189, 739–754.
- Hutchins, J.R., Toyoda, Y., Hegemann, B., Poser, I., Heriche, J.K., Sykora, M.M., Augsburg, M., Hudcz, O., Buschhorn, B.A., Bulkescher, J., et al. (2010). Systematic Analysis of Human Protein Complexes Identifies Chromosome Segregation Proteins. *Science* 328, 593–599.
- Kachur, T.M., Audhya, A., and Pilgrim, D.B. (2008). UNC-45 is required for NMY-2 contractile function in early embryonic polarity establishment and germline cellularization in *C. elegans*. *Dev. Biol.* 314, 287–299.
- Kao, H.L., and Gunsalus, K.C. (2008). Browsing multidimensional molecular networks with the generic network browser (N-Browse). *Curr. Protoc. Bioinformatics*, Chapter 9:Unit 9.11.
- Kimble, J., and Crittenden, S.L. Germline proliferation and its control (August 15, 2005), WormBook, ed. The *C. elegans* Research Community, WormBook, doi/10.1895/wormbook.1.13.1, <http://www.wormbook.org>.
- Kops, G.J., Voet, M.V., Manak, M.S., van Osch, M.H., Naini, S.M., Brear, A., McLeod, I.X., Hentschel, D.M., Yates, J.R., 3rd, van den Heuvel, S., et al. (2010). APC16 is a conserved subunit of the anaphase-promoting complex/cyclosome. *J. Cell Sci.* 123, 1623–1633.
- Lee, M.H., Ohmachi, M., Arur, S., Nayak, S., Francis, R., Church, D., Lambie, E., and Schedl, T. (2007). Multiple functions and dynamic activation of MPK-1 extracellular signal-regulated kinase signaling in *Caenorhabditis elegans* germline development. *Genetics* 177, 2039–2062.
- Li, S., Armstrong, C.M., Bertin, N., Ge, H., Milstein, S., Boxem, M., Vidalain, P.O., Han, J.D., Chesneau, A., Hao, T., et al. (2004). A map of the interactome network of the metazoan *C. elegans*. *Science* 303, 540–543.
- Liu, T., Sims, D., and Baum, B. (2009). Parallel RNAi screens across different cell lines identify generic and cell type-specific regulators of actin organization and cell morphology. *Genome Biol.* 10, R26.
- Maduro, M., and Pilgrim, D. (1995). Identification and cloning of *unc-119*, a gene expressed in the *Caenorhabditis elegans* nervous system. *Genetics* 141, 977–988.
- Mathey-Prevot, B., and Perrimon, N. (2006). *Drosophila* genome-wide RNAi screens: are they delivering the promise? *Cold Spring Harb. Symp. Quant. Biol.* 71, 141–148.
- Neumann, B., Walter, T., Heriche, J.K., Bulkescher, J., Erfle, H., Conrad, C., Rogers, P., Poser, I., Held, M., Liebel, U., et al. (2010). Phenotypic profiling of the human genome by time-lapse microscopy reveals cell division genes. *Nature* 464, 721–727.
- Piano, F., Gunsalus, K.C., Hill, D.E., and Vidal, M. (2006). *C. elegans* network biology: a beginning (August 21, 2006), WormBook, ed. The *C. elegans* Research Community, WormBook, doi/10.1895/wormbook.1.118.1, <http://www.wormbook.org>.
- Piano, F., Schetter, A.J., Mangone, M., Stein, L., and Kemphues, K.J. (2000). RNAi analysis of genes expressed in the ovary of *Caenorhabditis elegans*. *Curr. Biol.* 10, 1619–1622.
- Piano, F., Schetter, A.J., Morton, D.G., Gunsalus, K.C., Reinke, V., Kim, S.K., and Kemphues, K.J. (2002). Gene clustering based on RNAi phenotypes of ovary-enriched genes in *C. elegans*. *Curr. Biol.* 12, 1959–1964.
- Praitis, V., Casey, E., Collar, D., and Austin, J. (2001). Creation of low-copy integrated transgenic lines in *Caenorhabditis elegans*. *Genetics* 157, 1217–1226.
- Schuldiner, M., Collins, S.R., Thompson, N.J., Denic, V., Bhamidipati, A., Punna, T., Ihmels, J., Andrews, B., Boone, C., Greenblatt, J.F., et al. (2005). Exploration of the function and organization of the yeast early secretory pathway through an epistatic minarray profile. *Cell* 123, 507–519.
- Shi, A., Chen, C.C., Banerjee, R., Glodowski, D., Audhya, A., Rongo, C., and Grant, B.D. (2010). EHBP-1 functions with RAB-10 during endocytic recycling in *Caenorhabditis elegans*. *Mol. Biol. Cell* 21, 2930–2943.
- Sijen, T., Fleenor, J., Simmer, F., Thijssen, K.L., Parrish, S., Timmons, L., Plasterk, R.H., and Fire, A. (2001). On the role of RNA amplification in dsRNA-triggered gene silencing. *Cell* 107, 465–476.
- Simonis, N., Rual, J.F., Carvunis, A.R., Tasan, M., Lemmens, I., Hirozane-Kishikawa, T., Hao, T., Sahalie, J.M., Venkatesan, K., Gebreab, F., et al. (2009). Empirically controlled mapping of the *Caenorhabditis elegans* protein-protein interactome network. *Nat. Methods* 6, 47–54.
- Sönichsen, B., Koski, L.B., Walsh, A., Marschall, P., Neumann, B., Brehm, M., Alleaume, A.M., Artelt, J., Bettencourt, P., Cassin, E., et al. (2005). Full-genome RNAi profiling of early embryogenesis in *Caenorhabditis elegans*. *Nature* 434, 462–469.
- Starr, D.A., Hermann, G.J., Malone, C.J., Fixsen, W., Priess, J.R., Horvitz, H.R., and Han, M. (2001). *unc-83* encodes a novel component of the nuclear envelope and is essential for proper nuclear migration. *Development* 128, 5039–5050.
- Tatusov, R.L., Fedorova, N.D., Jackson, J.D., Jacobs, A.R., Kiryutin, B., Koonin, E.V., Krylov, D.M., Mazumder, R., Mekhedov, S.L., Nikolskaya, A.N., et al. (2003). The COG database: an updated version includes eukaryotes. *BMC Bioinformatics* 4, 41.
- Wilmes, G.M., Bergkessel, M., Bandyopadhyay, S., Shales, M., Braberg, H., Cagney, G., Collins, S.R., Whitworth, G.B., Kress, T.L., Weissman, J.S., et al. (2008). A genetic interaction map of RNA-processing factors reveals links between Sem1/Dss1-containing complexes and mRNA export and splicing. *Mol. Cell* 32, 735–746.
- Zhou, K., Rolls, M.M., Hall, D.H., Malone, C.J., and Hanna-Rose, W. (2009). A ZYG-12-dynein interaction at the nuclear envelope defines cytoskeletal architecture in the *C. elegans* gonad. *J. Cell Biol.* 186, 229–241.
- Zipperlen, P., Fraser, A.G., Kamath, R.S., Martinez-Campos, M., and Ahringer, J. (2001). Roles for 147 embryonic lethal genes on *C. elegans* chromosome I identified by RNA interference and video microscopy. *EMBO J.* 20, 3984–3992.

Supplemental Information

EXTENDED EXPERIMENTAL PROCEDURES

Target Selection

The sterile collection was defined as the set of 554 gene targets annotated as “sterile” in two or more prior studies. The studies that identified the most targets were Rual et al. (2004; 261 genes), Piano et al. (2002; 66 genes), Kamath et al. (2003; 258 genes), Fraser et al. (2000; 43 genes), Simmer et al. (2003; 298 genes), Maeda et al. (2001; 113 genes), and Sönnichsen et al. (2005; 139 genes). Smaller numbers of targets were included based on their identification in several additional studies (Chen et al., 2004; Choi et al., 2003; Davis et al., 2002; Detwiler et al., 2001; Fernandez et al., 2005; Fujita et al., 2002; Gönczy et al., 2000; Gottschalk et al., 2005; Hanazawa et al., 2001; Kaitna et al., 2002; Kinnaird et al., 2004; Maciejowski et al., 2005; Marin and Evans, 2003; Martin et al., 2002; Molin and Puisieux, 2005; Olgun et al., 2005; Skop et al., 2004; Xu et al., 2005).

Oligo Design

Templates for dsRNA production were generated for each target by using PCR to amplify a 500-1000 bp region of the corresponding gene. One oligo in each pair had a tail containing the T3 RNA polymerase promoter and the other had a tail containing the T7 promoter. When possible, oligo pairs corresponding to those used previously by Sönnichsen et al. (2005) were chosen. New oligo pairs were designed for genes not in the Sönnichsen screen and for cases in which the Sönnichsen oligo pairs either amplified introns or regions smaller or larger than 500-1000 bp. A comprehensive list of the oligo pairs used in this screen can be found in the Supplemental excel workbook **Table S2**. Oligo sequences can also be accessed, along with all of the movies and phenotypic data from the screen, via the Phenobank website (http://worm.mpi-cbg.de/phenobank_gonad).

PCR

50 μ L PCR reactions were set up to final concentrations as follows: 0.3 mM each dNTP, 1xTaq Polymerase buffer, 0.01 mg/ml N2 genomic DNA template (0.5 μ L of 1mg/mL), 50 U/mL Taq Polymerase (0.5 μ L of 5000 U/mL), and 1 mM of each oligo (1 μ L of a mixture of forward and reverse oligos at 50 mM each). Reactions were prepared by aliquoting 49 μ L of a master mix containing all components except the oligos into each well and subsequently adding the oligo mix using a multichannel pipet. Thermocycling parameters were as follows: 92°C, 5 min, 34 x [92°C, 20 s; 58°C, 40 s; 72°C, 90 s], 72°C, 5 min, 4°C hold. 5 μ L of each PCR reaction was analyzed following electrophoresis on a 1% agarose gel (TBE). Each reaction was scored for the presence of bands of the appropriate size and intensity and the absence of non-specific products. Oligos were redesigned for thirteen targets that we were unable to troubleshoot. Successful PCR products were cleaned up using the PerfectPrep 96 kit (Eppendorf) according to manufacturer’s recommendations.

RNA Transcription and Annealing

Transcription reactions (16 hr) and DNase treatment (20 min) were performed at 37°C in 96 well plates in a thermocycler according to manufacturer’s recommendations (Megascript T3 and T7 kits, Ambion). The T3 and T7 reactions for each template were mixed and cleaned (MegaClear96 kit, Ambion) according to manufacturers recommendations. The T3 and T7 strands were annealed by adding 3x Soaking buffer (32.7 mM Na₂HPO₄, 16.5 mM KH₂PO₄, 6.3 mM NaCl, 14.1 mM NH₄Cl) to a final concentration of 1X and incubating the reactions at 68°C for 10 min followed by 37°C for 30 min. dsRNAs were partitioned into 5 μ L aliquots and stored at –80°C until use. In the 21 cases where soaking in dsRNA gave rise to worms that were not sterile and had no detectable gonad morphology defects, dsRNAs were re-made and the experiment was repeated.

Soaking RNAi

There are three methods for dsRNA delivery in *C. elegans*: injection, soaking and feeding. In the context of the gonad, feeding is the weakest and most variable delivery method. In our hands injection and soaking are both highly reproducible and give comparable effects, with soaking in some cases being superior (i.e., Maddox et al., 2005). Since soaking is technically easier and more reproducible, we chose to deliver dsRNAs by soaking. Larval (L4) stage worms from the strain OD95 (genotype: *unc-119(ed3)* III; *lts37[pAA64]; pie-1/mCherry::his-58; unc-119(+)* IV; *lts38[pAA1;pie-1/GFP::PH(PLC1 delta1); unc-119(+)]*), expressing the plasma membrane marker (GFP-PH(PLC1 δ 1)) and the chromosome marker (mCherry-Histone H2B) were rinsed with M9 to clean off residual bacteria and soaked in 5 μ L of dsRNA (supplemented with 0.5 μ L of a 50/50 mixture of 63 mM spermidine and 1.1% gelatin) for 24 hr at 20°C inside a humidifying chamber. After soaking, the worms were transferred to NGM plates seeded with OP50 *E. coli* and returned to the 20°C incubator to recover for 48 hr prior to imaging.

Gonad Imaging

The reproductive organ, or gonad, of the adult *C. elegans* hermaphrodite (Figure 1C) is a dynamic syncytium that houses ~1000 nuclei in cup-shaped compartments open to a common cytoplasmic core, called the “rachis.” The compartments mature into oocytes as they progress from the distal tip to the proximal region of the gonad adjacent to the spermatheca. Thus, the temporal stages of oocyte development are laid out in space within the gonad. Nuclei are born in mitotic divisions at the distal tip of the gonad and, after replicating their DNA, progress through meiotic prophase as they move from the distal region toward the proximal end of the gonad. Approximately half of the nuclei are culled by apoptosis prior to the gonad turn (filled green circles prior to the turn in the

schematic in Figure 1C are apoptotic nuclei). As compartments round the turn, they expand dramatically in size to generate the nascent oocytes. This expansion is fueled by transcription from nuclei in the pachytene region of the gonad, which serve as “nurse” nuclei that help generate the large quantities of protein, membranes, and mRNA that are loaded into the developing oocytes. After rounding the turn, the oocytes bud off of the rachis tip. Typically the 3–4 oocytes immediately adjacent to the spermatheca have budded off from the common cytoplasm. The oocytes are fertilized as they pass through the spermatheca at the proximal end of the gonad. In control worms, each of the two gonad “arms” generates one fertilized oocyte every 20 min.

Gonads were imaged by anesthetizing six living worms using a fresh mixture of 1 mg/ml Tricane (ethyl 3-aminobenzoate methanesulfonate salt) and 0.1 mg/ml of tetramisole hydrochloride (TMHC) dissolved in M9 for 15–30 min. Anesthetized worms were transferred to an agarose pad, covered with a coverslip and immediately imaged. Microscopy was performed with a spinning disk confocal mounted on a Nikon TE2000-E inverted microscope equipped with a 60x 1.4 NA PlanApochromat lens, a krypton-argon 2.5 W water-cooled laser (Spectra-Physics, Mountain View, CA) and a Hamamatsu Orca ER CCD camera. Acquisition parameters, shutters, and focus were controlled by MetaMorph software (Molecular Devices, Downingtown, PA). Three gonads were imaged for each depletion. For each gonad, an 80 × 0.5 mm z-series was collected including DIC, GFP and RFP images for every z-plane.

Sterility Assessment

To assess sterility resulting from RNA treatment, 3 worms that had been soaked in dsRNA for 24 hr and allowed to recover for 48 hr were moved to a fresh plate at the time when the other worms in the treated cohort were anesthetized and imaged. After 24 hr at 20°C, the number of embryos and L1 progeny on the plate was counted. Three wild-type worms will lay ~150 embryos over a 24 hr period, and the embryos laid during the first 12 hr will have hatched to generate L1 larvae. A gene was designated as “sterile” if its knockdown yielded plates with no embryos or L1 larvae. If < 50 embryos/L1 worms (<33% of the expected wild-type brood size) were present, the gene was designated “partially sterile” and if > 50 embryos/L1 worms were present the gene was considered “wild-type” for sterility (Figure S1B). 24 hr after the embryos/L1 larvae were counted, we assessed whether gene knockdown resulted in embryonic lethality or viable progeny. If > 50 L1–L2 larvae were present, they were considered “viable,” if < 50 L1–L2 larvae were present, plates were inspected for the presence of dead embryos; if there were > 5 dead embryos, it was considered “embryonic lethal.” The breakdown of genes by sterility class is summarized in Figure S1B. We note that many of the gene knockdowns scored as “wild-type” for brood size, nevertheless exhibited detectable gonad morphology phenotypes (of the 107 knockdowns whose brood size was classified as “wild-type,” yellow pie slice, 85 gave a gonad phenotype and 22 had wild-type gonad morphology). This apparent discrepancy is likely due to the fact that to be classified as “partially sterile” a knockdown had to reduce brood size by more than 2/3. Thus, many of the knockdowns classified as “wild-type” do reduce brood size, but to a lesser extent. If a gene knockdown resulted in > 50 viable progeny in the sterility test and no gonad phenotype was observed (assessed as described in the next section), the experiment was repeated. If the result was the same, new oligos were designed for the targeted gene and the experiment was repeated a second time.

Data Annotation and Placement into Phenotypic Classes

During the classification process, movie sets were indexed by RNA number so they could be analyzed blind to the identity of the gene targets. Binary phenotypic profiles were generated for each gene knockdown by scoring the set of 3 movies for the presence or absence (1 = defect present, 0 = defect absent) of 94 possible defects (for a complete list of defects including examples of each, see Table S1). Not all defects were independent. To reflect increasing severity of the same phenotypic feature, some defects were scored in a cumulative fashion (see cumulative defects scored table below). For example, reduced apoptosis-69 was scored by assigning “1” to defect 69 and the more severe phenotype, absent apoptosis-70, was scored by assigning “1” to both defects 69 and 70.

During annotation, the set of three movies acquired for each knockdown was viewed by the same pair of individuals, who viewed them together and came to a consensus with respect to each phenotypic feature. For each of the 94 defects, a “1” was assigned when the defect was present in at least 2 of the 3 movies. For phenotypic features in which cumulative sets of defects were used to reflect increasing severity, the most severe defect designator was assigned when present in one sequence if the less severe version of the same feature was present in one or both of the other two sequences. For example, if no oocytes were formed in one movie and one or both of the other two movies exhibited reduced oocyte number, the movie set would be annotated with both no oocytes formed-67 and decreased oocyte number- 61. In the 24 cases where the three movies exhibited different phenotypes, the experiment was repeated. If a data set was scored as wild-type and the sterility assessment was wild-type, new oligos were designed and the experiment was repeated.

Calculation of the Connection Specificity Index

For each pair of genes A and B, CSI_{AB} was calculated by: (1) calculating the PCC for the connections between A and B and each of the other genes in the dataset, (2) counting the number of genes whose profiles correlated with those of A or B at a level comparable to or better than the correlation between A and B (illustrated in Figure 4C; correlations with a PCC up to 0.05 less than PCC_{AB} were considered similar—this offset was determined empirically); 3) dividing this number by the total number of genes in the screen (554); and 4) subtracting the result from 1.0. Generating CSI scores for the Sönnichsen et al. (2005) embryo timelapse data required binarizing it. Originally, this data set was not binary; instead, for each of the 45 possible phenotypic features, the prevalence (fraction of embryos in

which the feature was observed; i.e if a feature was observed in 2 out of 5 embryos, the prevalence was 2/5) was scored. This data was binarized by assigning a “1” in cases in which the feature was observed in 2 or more embryos and a “0” in cases in which the feature was observed in 0 or 1 embryos.

SUPPLEMENTAL REFERENCES

- Chen, Z., Eastburn, D.J., and Han, M. (2004). The *Caenorhabditis elegans* nuclear receptor gene nhr-25 regulates epidermal cell development. *Mol. Cell. Biol.* 24, 7345–7358.
- Choi, K.Y., Ji, Y.J., Dhakal, B.K., Yu, J.R., Cho, C., Song, W.K., and Ahnn, J. (2003). Vacuolar-type H⁺-ATPase E subunit is required for embryogenesis and yolk transfer in *Caenorhabditis elegans*. *Gene* 311, 13–23.
- Davis, E.S., Wille, L., Chestnut, B.A., Sadler, P.L., Shakes, D.C., and Golden, A. (2002). Multiple subunits of the *Caenorhabditis elegans* anaphase-promoting complex are required for chromosome segregation during meiosis I. *Genetics* 160, 805–813.
- Detwiler, M.R., Reuben, M., Li, X., Rogers, E., and Lin, R. (2001). Two zinc finger proteins, OMA-1 and OMA-2, are redundantly required for oocyte maturation in *C. elegans*. *Dev. Cell* 1, 187–199.
- Fernandez, A.G., Gunsalus, K.C., Huang, J., Chuang, L.S., Ying, N., Liang, H.L., Tang, C., Schetter, A.J., Zegar, C., Rual, J.F., et al. (2005). New genes with roles in the *C. elegans* embryo revealed using RNAi of ovary-enriched ORFeome clones. *Genome Res.* 15, 250–259.
- Fujita, M., Takasaki, T., Nakajima, N., Kawano, T., Shimura, Y., and Sakamoto, H. (2002). MRG-1, a mortality factor-related chromodomain protein, is required maternally for primordial germ cells to initiate mitotic proliferation in *C. elegans*. *Mech. Dev.* 114, 61–69.
- Gottschalk, A., Almedom, R.B., Schedletzky, T., Anderson, S.D., Yates, J.R., 3rd, and Schafer, W.R. (2005). Identification and characterization of novel nicotinic receptor-associated proteins in *Caenorhabditis elegans*. *EMBO J.* 24, 2566–2578.
- Hanazawa, M., Mochii, M., Ueno, N., Kohara, Y., and Iino, Y. (2001). Use of cDNA subtraction and RNA interference screens in combination reveals genes required for germ-line development in *Caenorhabditis elegans*. *Proc. Natl. Acad. Sci. USA* 98, 8686–8691.
- Kaitra, S., Pasierbek, P., Jantsch, M., Loidl, J., and Glotzer, M. (2002). The aurora B kinase AIR-2 regulates kinetochores during mitosis and is required for separation of homologous Chromosomes during meiosis. *Curr. Biol.* 12, 798–812.
- Kamath, R.S., Fraser, A.G., Dong, Y., Poulin, G., Durbin, R., Gotta, M., Kanapin, A., Le Bot, N., Moreno, S., Sohrmann, M., et al. (2003). Systematic functional analysis of the *Caenorhabditis elegans* genome using RNAi. *Nature* 421, 231–237.
- Kinnauld, J.H., Maitland, K., Walker, G.A., Wheatley, I., Thompson, F.J., and Devaney, E. (2004). HRP-2, a heterogeneous nuclear ribonucleoprotein, is essential for embryogenesis and oogenesis in *Caenorhabditis elegans*. *Exp. Cell Res.* 298, 418–430.
- Kachur, T.M., Audhya, A., and Pilgrim, D.B. (2008). UNC-45 is required for NMY-2 contractile function in early embryonic polarity establishment and germline cellularization in *C. elegans*. *Dev. Biol.* 314, 287–299.
- Maciejowski, J., Ahn, J.H., Cipriani, P.G., Kilian, D.J., Chaudhary, A.L., Lee, J.I., Voutev, R., Johnsen, R.C., Baillie, D.L., Gunsalus, K.C., et al. (2005). Autosomal genes of autosomal/X-linked duplicated gene pairs and germ-line proliferation in *Caenorhabditis elegans*. *Genetics* 169, 1997–2011.
- Maeda, I., Kohara, Y., Yamamoto, M., and Sugimoto, A. (2001). Large-scale analysis of gene function in *Caenorhabditis elegans* by high-throughput RNAi. *Curr. Biol.* 11, 171–176.
- Marin, V.A., and Evans, T.C. (2003). Translational repression of a *C. elegans* Notch mRNA by the STAR/KH domain protein GLD-1. *Development* 130, 2623–2632.
- Martin, E., Laloux, H., Couette, G., Alvarez, T., Bessou, C., Hauser, O., Sookhareea, S., Labouesse, M., and Segalat, L. (2002). Identification of 1088 new transposon insertions of *Caenorhabditis elegans*: a pilot study toward large-scale screens. *Genetics* 162, 521–524.
- Molin, L., and Puisieux, A. (2005). *C. elegans* homolog of the Caf1 gene, which encodes a subunit of the CCR4-NOT complex, is essential for embryonic and larval development and for meiotic progression. *Gene* 358, 73–81.
- Olgun, A., Alekseenko, T., Pereira-Smith, O.M., and Vassilatis, D.K. (2005). Functional analysis of MRG-1: the ortholog of human MRG15 in *Caenorhabditis elegans*. *J. Gerontol. A Biol. Sci. Med. Sci.* 60, 543–548.
- Rual, J.F., Ceron, J., Koreth, J., Hao, T., Nicot, A.S., Hirozane-Kishikawa, T., Vandenhoute, J., Orkin, S.H., Hill, D.E., van den Heuvel, S., et al. (2004). Toward improving *Caenorhabditis elegans* phenome mapping with an ORFeome-based RNAi library. *Genome Res.* 14, 2162–2168.
- Simmer, F., Moorman, C., van der Linden, A.M., Kuijk, E., van den Berghe, P.V., Kamath, R.S., Fraser, A.G., Ahringer, J., and Plasterk, R.H. (2003). Genome-wide RNAi of *C. elegans* using the hypersensitive rrf-3 strain reveals novel gene functions. *PLoS Biol.* 1, E12.
- Skop, A.R., Liu, H., Yates, J., 3rd, Meyer, B.J., and Heald, R. (2004). Dissection of the mammalian midbody proteome reveals conserved cytokinesis mechanisms. *Science* 305, 61–66.
- Xu, X., Lee, D., Shih, H.Y., Seo, S., Ahn, J., and Lee, M. (2005). Linking integrin to IP(3) signaling is important for ovulation in *Caenorhabditis elegans*. *FEBS Lett.* 579, 549–553.

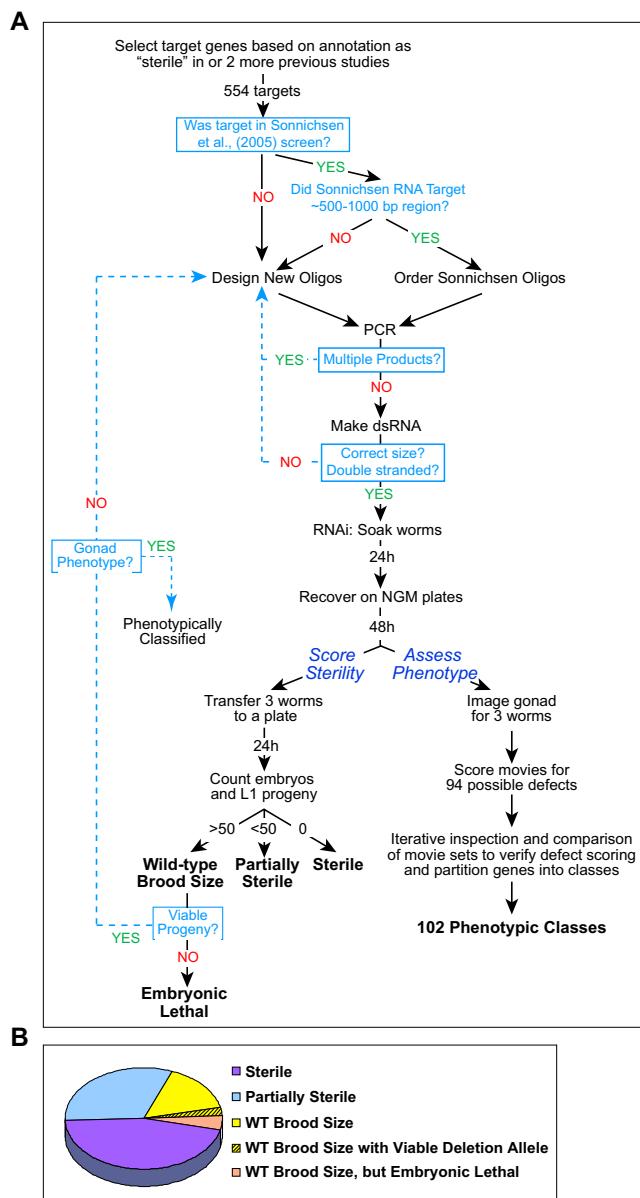


Figure S1. Design and Execution of the High-Content Gonad Morphology Screen and Sterility Assessment, Related to Figure 1
(A) Flow chart summarizes RNA production, RNAi, fluorescence imaging protocol, and assessment of sterility.
(B) Pie chart summarizes sterility results for the 554 genes addressed in this study.

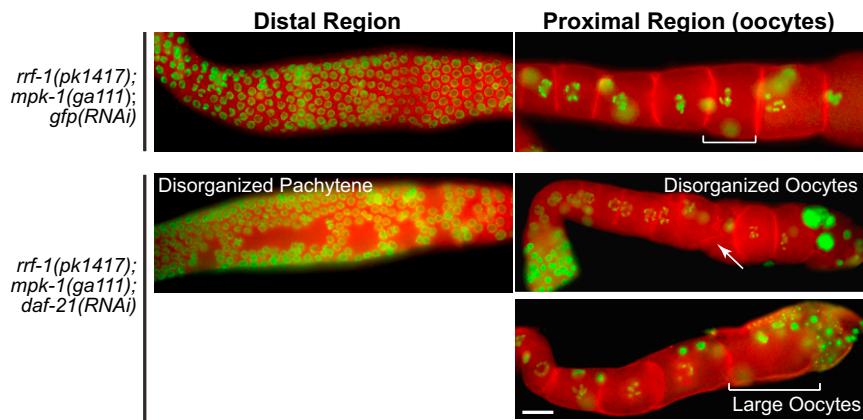


Figure S2. Phenotypes Scored in the *mpk-1(ga111)* Enhancement Assay, related to Figure 2

The *mpk-1* partial loss-of-function allele *ga111* leads to an amino acid substitution near the activation loop of MPK-1 that causes a temperature-sensitive defect in its ability to be activated by MEK-2. At the permissive temperature (20°), *mpk-1(ga111)* gonads have essentially normal morphology, although levels of phosphorylated MPK-1 (dpMPK-1), a readout for pathway activity, are reduced compared to controls (Figure 2E,F, Table S3; Lee et al., 2007). To specifically examine the germline function of the MAPK pathway rather than its role in the surrounding somatic tissues, experiments were performed in an *rrf-1(pk1417)* null mutant background. *rrf-1* encodes an RNA-dependent RNA polymerase that is required for RNAi in somatic cells but not in the germline (Sijen et al., 2001). Gonads were fixed and stained to visualize the plasma membrane (antibodies to SYN-4 and PTC-1, red), and chromosomes (DAPI, green). Images show the distal and proximal regions of control (*gfp(RNAi)*; **top row**) and *daf-21(RNAi)*; **middle and bottom rows**) gonads in the *rrf-1(pk1417) mpk-1(ga111)* background. Partial RNAi of *daf-21* in the *mpk-1(ga111)* background can result in disorganized pachytene nuclei or oocytes, and/or irregular-sized oocytes. Scale bar is 10 μ m.

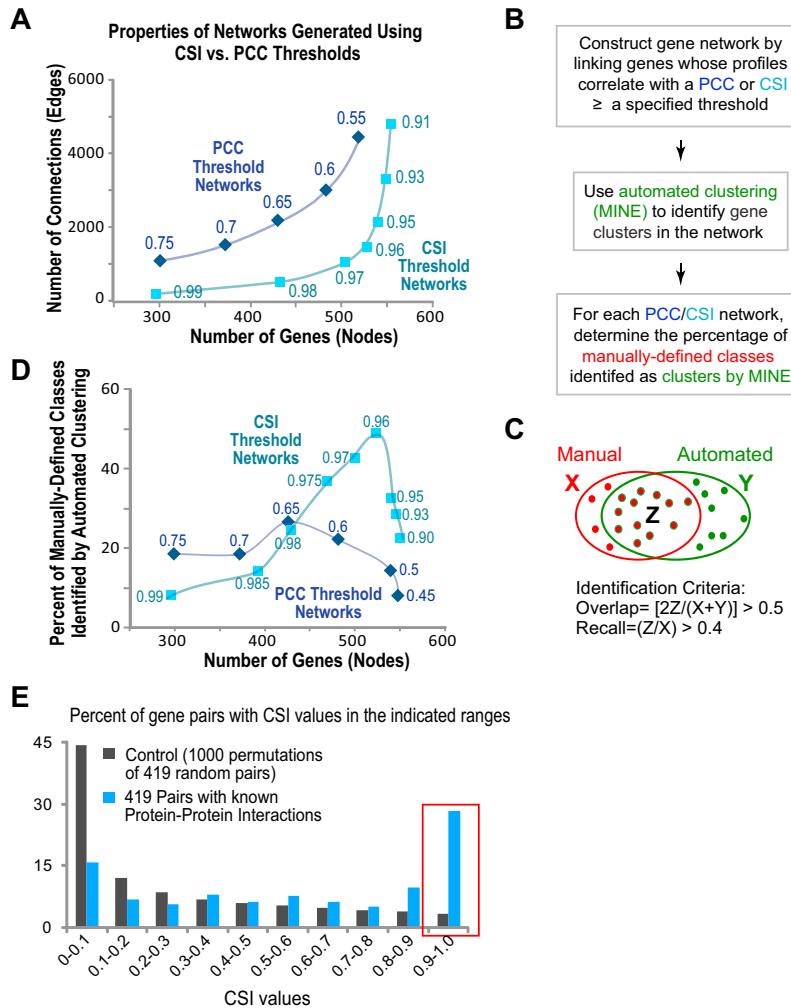


Figure S3. Comparison of PCC and CSI Based Gene Networks, Related to Figure 4

To evaluate the effectiveness of the CSI, we began by comparing the properties of gene networks constructed by displaying connections with PCC or CSI values greater than or equal to specific thresholds. (A) Plot showing the properties of networks generated using the indicated PCC (dark blue) or CSI (cyan) thresholds. The PCC or CSI threshold used to generate the network is indicated next to each data point; the x and y axis values for each point correspond to the number of genes (**nodes**) and the number of connections (**edges**) in the network generated using that threshold. As the CSI/PCC thresholds are increased, increasing numbers of genes drop out of the network because they do not make any connections with PCC/CSI values higher than the threshold. Over the useful range of the two parameters, networks generated using CSI thresholds have substantially (~3 fold) fewer connections than PCC networks that include comparable numbers of genes. (B) Flowchart outlining how the performance of CSI and PCC threshold networks was evaluated by comparing the ability of an automated clustering algorithm ("Module Identification in Networks" or MINE; K. Rhrissorakrai and KCG, *in press*) to identify gene clusters corresponding to the 49 manually-defined phenotypic classes containing 4 or more genes. MINE iteratively identifies clusters of genes based on network topology and employs a small set of user-defined parameters for modularity, node degree, and clustering coefficient (the same set of parameters was used for all analyses). During a first round of partitioning, nodes are allowed to be in multiple clusters. After this initial round, clusters were merged if the overlap fraction between them was > 0.75 . (C) Schematic of the method used to determine whether a manually-defined phenotypic class was identified by the automated clustering algorithm. A manual cluster was considered identified if the overlap between the computational cluster and the manual cluster was > 0.5 and the fraction of the genes in the manual cluster recovered by the computational cluster (recall) was > 0.4 . (D) Plot showing the percent of manually-defined phenotypic classes that could be identified by automated clustering for networks generated using the indicated PCC (dark blue) and CSI (cyan) thresholds. At the single best CSI threshold of 0.96, MINE was able to identify 49% of the manually-defined classes containing 4 or more genes; 88% of manually-defined classes could be identified when the algorithm was allowed to search networks that spanned the relevant range of CSI thresholds (0.90 to 0.99), reflecting the fact that our manually-defined phenotypic classes do not correspond to a uniform level of functional resolution. For the PCC networks, MINE was able to identify 27% of the manually-defined classes at the best PCC threshold of 0.65, and 65% of manually-defined classes could be identified when the algorithm was allowed to search networks that spanned the relevant range of PCC thresholds (0.5-1). The superior performance of CSI networks in this analysis, suggests that the CSI reduces edge noise and increases network clarity. (E) Protein-protein interactions also provide evidence for shared biological function and thus overlap with phenotypic links more than expected by chance. Thus, to identify the range of potentially meaningful CSI values in the sterile network, we also analyzed the CSI distribution for the protein pairs in the sterile dataset that were also previously shown to physically interact by 2-hybrid. Among the 554 genes in the sterile collection, 419 pairs shared a physical connection in a composite protein interaction map (Boxem et al., 2008; Li et al., 2004; Simonis et al., 2009); these pairs have a mean CSI = 0.57 and standard deviation (SD) = 0.35. For comparison,

a background distribution was calculated based on 1000 random samples of 419 gene pairs from the sterile collection set (CSI mean = 0.26, SD = 0.29). Gaussian Kernel densities were plotted for both distributions. The distribution of CSI scores for the 419 gene pairs with protein-protein interactions is significantly different from the distribution obtained for random controls, being heavily skewed toward CSI values > 0.90 (**red box**).

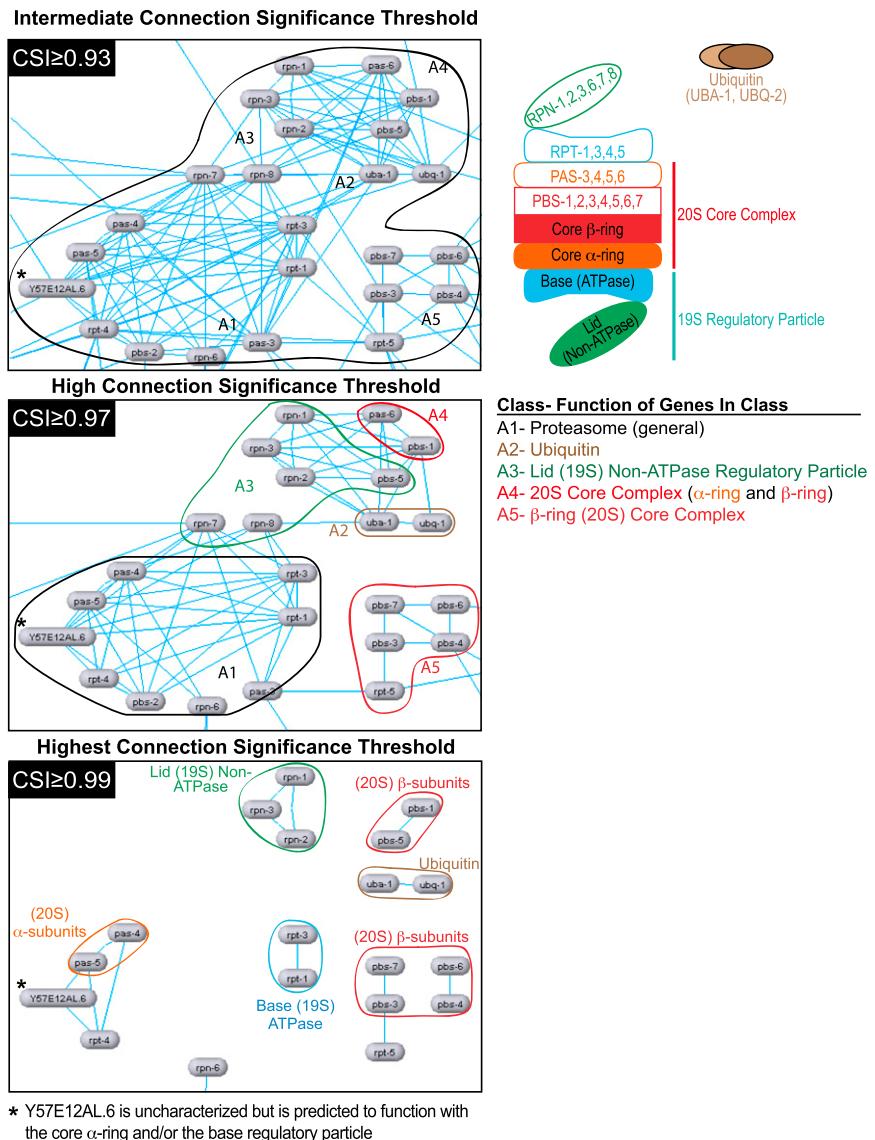


Figure S4. Profiling gene knockdowns in the gonad yields a functional map with a level of resolution sufficient to place components in specific proteasome subcomplexes, related to Figure 5

The region of the gene network involved in protein degradation is shown using three different CSI thresholds (0.93, 0.97 and 0.99, as indicated) to filter the displayed gene connections. Gene groups from different manually-assigned classes are circled in the middle panel (CSI ≥ 0.97). Gene groups corresponding to specific proteasome subcomplexes are circled in the lower panel (CSI ≥ 0.99). Diagram illustrates the sub-molecular complexes of the proteasome and putative location of the *C. elegans* proteasome proteins shown in the network. Using the very high CSI threshold of 0.99, the connections that remain link genes within specific proteasome subcomplexes (the Lid, the core β -ring, the core α -ring or the 19S ATPase base). One uncharacterized gene, Y57E12AL.6, maintains connections with the 20S α -subunits at this high (0.99) level, suggesting that it may be a component of the α -ring. The phenotypic resolution that we obtain for the proteasome, based on gonad morphology, compares favorably with that obtained based on genetic interaction profiling in budding yeast (Breslow et al., 2008) indicating that phenotypic analysis exploiting the biological complexity of the gonad, coupled with CSI-based automated analysis of the phenotypic parameter dataset, is capable of correctly assigning very fine distinctions in protein function.

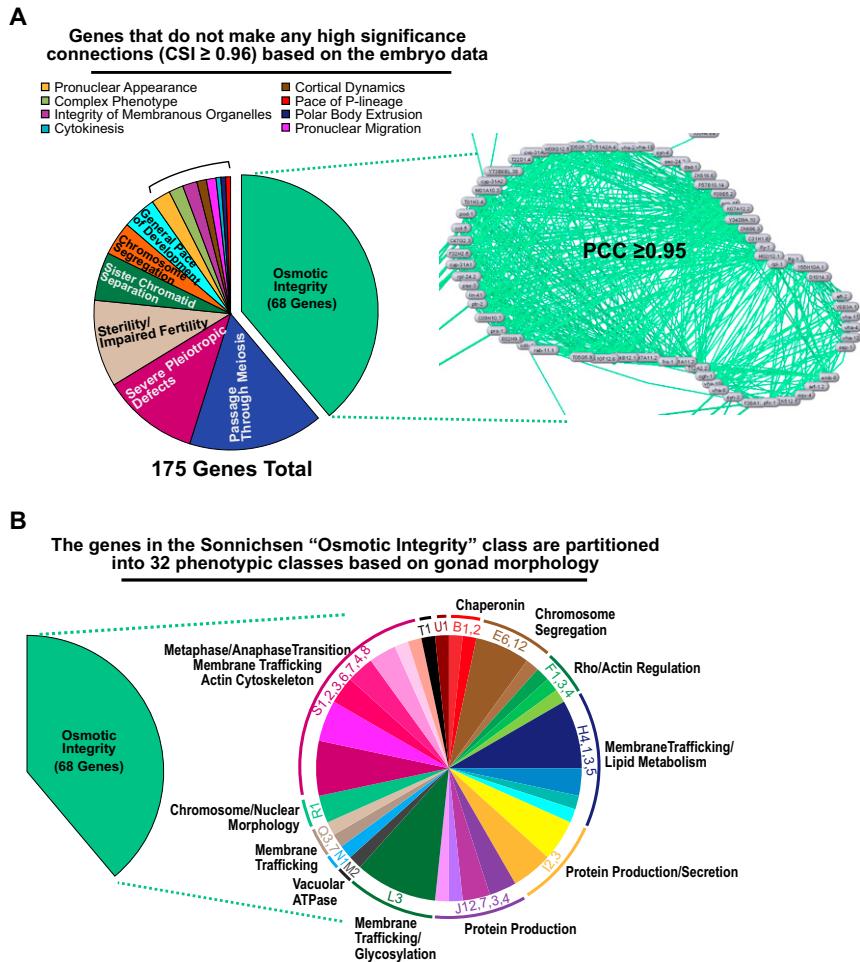


Figure S5. The CSI framework allows data set integration by filtering out connections based on low information content phenotypes, related to Figure 6

Cumulatively, the analysis of the sterile collection based on gonad morphology and the timelapse DIC filming of embryos (Sönnichsen, 2005) provide knockdown profiles for 885 genes. Attempts to generate an integrated network based on the Pearson's Correlation Coefficient fail because the embryo data groups the majority of the 330 genes profiled in both screens into large interconnected clusters. This occurs because genes required for embryo production exhibit very similar (highly-correlated) phenotypes in the embryo-filming assay that provide little resolving power because they are shared by many genes. To integrate the two datasets, we chose a CSI threshold of 0.96 to visualize the combined network because empirical tests indicated that higher thresholds eliminated meaningful connections from the network and lower thresholds resulted in non-specific noise. The integrated network contains 818 genes (or ~90% of the total number of tested genes) that make at least one high-specificity connection ($CSI \geq 0.96$; Figure 6). Using the CSI to filter out non-specific connections, we found that 138 of the 330 genes in both datasets made no informative connections based on the embryo-filming data, and thus remained essentially uncharacterized. (A) Pie chart showing how the 175 genes for which the embryo data fails to identify any high significance ($CSI \geq 0.96$) connections are partitioned among the Sönnichsen phenotypic classes. These genes were primarily derived from four classes in the Sönnichsen screen "Osmotic Integrity," "Passage Through Meiosis," "Severe Pleiotropic Defects," "Sterility/Impaired Fertility." The "Osmotic Integrity" class is an example of a highly-correlated but low specificity embryo phenotype. The 68 genes in the "Osmotic Integrity" class have phenotypic profiles that are highly correlated; the "hairball" to the right shows the genes in the Osmotic Integrity class displaying only connections above a PCC threshold of 0.95. Despite the high level of correlation, inhibition of many different genes/pathways can result in embryos that rupture due to osmotic stress, hence this phenotype provides little specific information on gene function. (B) Analysis of gonad morphology places the 68 genes in the Sönnichsen "Osmotic Integrity" class into 32 different phenotypic classes that make strong predictions for molecular function. Individual pie slices represent phenotypic classes (as labeled), while color spectrums within a hue represent the broad categories that encompass these phenotypic classes. Predicted molecular functions for the phenotypic classes are listed adjacent to the broad category groupings.