RESOURCE

nature cell biology

A combined binary interaction and phenotypic map of *C. elegans* cell polarity proteins

Thijs Koorman[1], Diana Klompstra[2,3,6], Monique van der Voet[1,6,7], Irma Lemmens[4,5], João J. Ramalho[1], Susan Nieuwenhuize[1], Sander van den Heuvel[1], Jan Tavernier[4,5], Jeremy Nance[2,3] and Mike Boxem[1,8]

The establishment of cell polarity is an essential process for the development of multicellular organisms and the functioning of cells and tissues. Here, we combine large-scale protein interaction mapping with systematic phenotypic profiling to study the network of physical interactions that underlies polarity establishment and maintenance in the nematode *Caenorhabditis elegans*. Using a fragment-based yeast two-hybrid strategy, we identified 439 interactions between 296 proteins, as well as the protein regions that mediate these interactions. Phenotypic profiling of the network resulted in the identification of 100 physically interacting protein pairs for which RNAi-mediated depletion caused a defect in the same polarity-related process. We demonstrate the predictive capabilities of the network by showing that the physical interaction between the RhoGAP PAC-1 and PAR-6 is required for radial polarization of the *C. elegans* embryo. Our network represents a valuable resource of candidate interactions that can be used to further our insight into cell polarization.

The ability to polarize is a fundamental cellular property, required for processes such as cell migration and asymmetric cell division, and for the specification of functionally distinct domains. A series of cortically localized proteins has been identified that can drive the establishment of cell polarity. The PAR-3–PAR-6–aPKC and Crumbs–SDT–PATJ complexes together promote apical domain identity[1–3], whereas the LGL–SCRIB–DLG proteins, originally identified as *Drosophila* tumour suppressors[4–6], promote basolateral identity[2]. These cortical polarity complexes act together with a number of other components, such as Rho-family GTPases, junctional components, and cytoskeletal linkers of the ERM family, in establishing polarity.

Although it is clear that mutual exclusion is a key mechanism by which cortical polarity regulators establish polarity[7,8], we still lack a detailed mechanistic understanding of how cortical polarity regulators are segregated into distinct domains. Moreover, we know little of the mechanisms through which cortical polarity is integrated with cellular events such as cytoskeletal rearrangement, organization of a polarized trafficking machinery, and functional specialization of membrane domains. A full understanding of polarity establishment will require a comprehensive knowledge of the proteins involved in this process and the molecular interactions between them.

Here, we study the network of physical interactions that underlies polarity establishment in the nematode *Caenorhabditis elegans* using a combination of large-scale yeast two-hybrid (Y2H) screens and phenotypic profiling. We identified a polarity interaction network of 439 interactions, and mapped the protein regions mediating these interactions. Phenotypic profiling by RNA-mediated interference (RNAi) revealed 100 protein pairs that exhibited a phenotype in the same polarity-related process. These pairs are strong candidates for a functional interaction *in vivo*. We studied the interaction between PAR-6 and PAC-1 ARHGAP21 in detail, and demonstrate that this physical interaction is important for radial polarization of the *C. elegans* embryo. Our data provide a resource for future studies into cell polarity, and should contribute to our understanding of this essential process. A searchable web interface of all interactions and fragments identified is available at http://www.projects.science.uu.nl/interactome.

RESULTS

Identification of the *C. elegans* polarity interaction network

To generate a map of interactions underlying polarity establishment in *C. elegans*, we selected 69 proteins that control cell polarity in

[1]Division of Developmental Biology, Department of Biology, Faculty of Science, Utrecht University, 3584 CH, Utrecht, The Netherlands. [2]Helen L. and Martin S. Kimmel Center for Biology and Medicine at the Skirball Institute of Biomolecular Medicine, NYU School of Medicine, New York, New York 10016, USA. [3]Department of Cell Biology, NYU School of Medicine, New York, New York 10016, USA. [4]Department of Medical Protein Research, VIB, 9000 Ghent, Belgium. [5]Department of Biochemistry, Faculty of Medicine and Health Sciences, Ghent University, 9000 Ghent, Belgium. [6]These authors contributed equally to this work. [7]Present address: Department of Human Genetics, Donders Institute for Brain, Cognition and Behaviour, Radboud University Medical Center, 6500 HB Nijmegen, The Netherlands. [8]Correspondence should be addressed to M.B. (e-mail: m.boxem@uu.nl)

Received 3 March 2015; accepted 15 December 2015; published online 18 January 2016; DOI: 10.1038/ncb3300

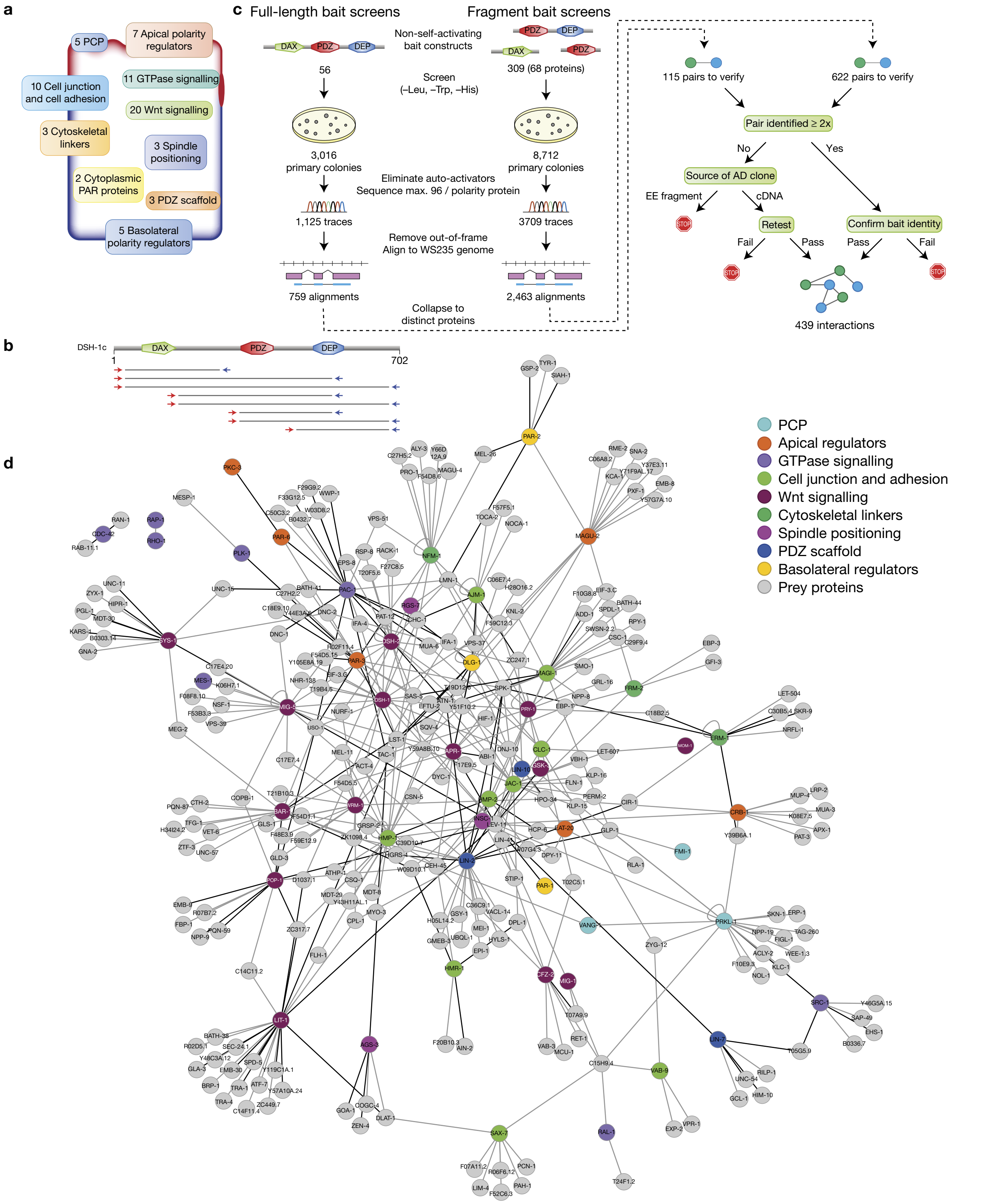


Figure 1 Identification and validation of a *C. elegans* polarity interaction network (*Ce*PIN). (**a**) Schematic representation of the different classes of protein selected as baits. Numbers indicate number of proteins in each class. See Supplementary Table 1 for details. (**b**) Example of the design of bait fragments. Fragments were manually designed on the basis of the size of the gene and the predicted protein domain composition. Protein domains are indicated. Grey lines represent the fragments cloned. Red and blue arrows represent forward and reverse amplification primers, respectively. (**c**) Overview of the Y2H screening and computational validation pipeline used to establish the *Ce*PIN. (**d**) Network graph of the identified protein interactions. Coloured nodes are bait proteins, and the colour indicates the classification in **a**. Grey nodes are prey proteins. Black edges indicate interactions with a phenotype overlap (see main text).

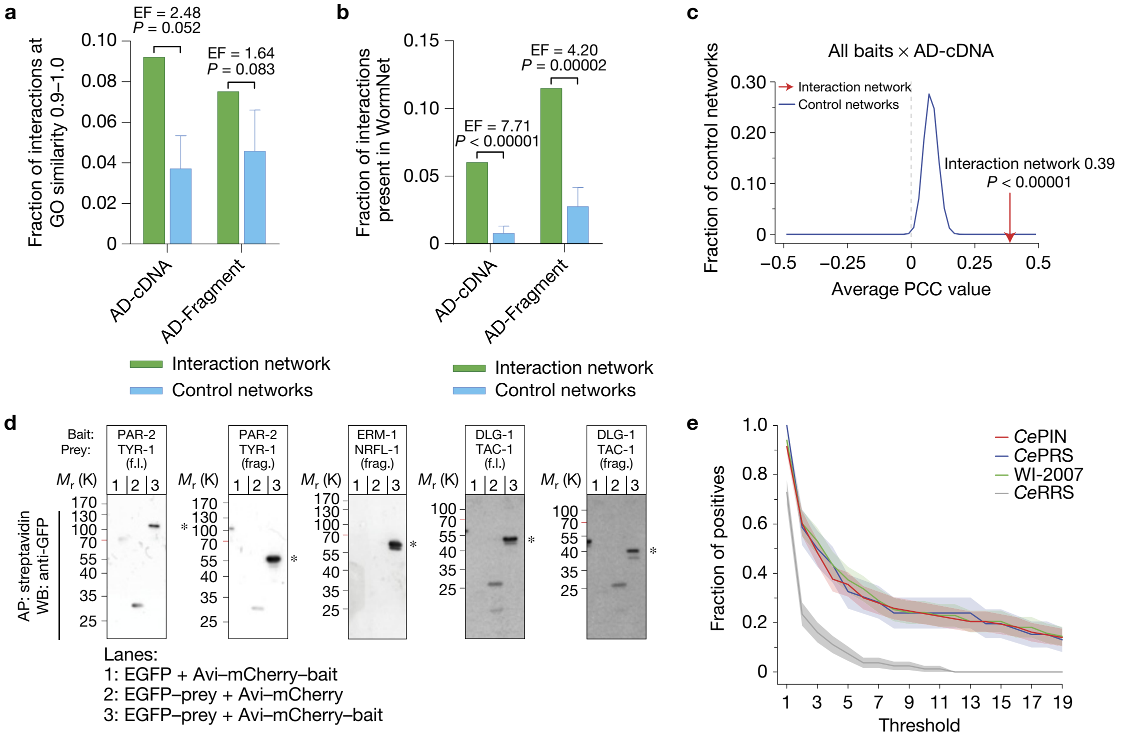


Figure 2 Validation of the *Ce*PIN. (**a**,**b**) Enrichment of similar Gene Ontology (GO) terms (**a**) and presence in WormNet (**b**) for interacting protein pairs compared with control networks generated by replacing prey proteins with random proteins from the search space. Interactions from AD-cDNA and AD-Fragment libraries were plotted separately, as the search space was very different. Control network bars represent the mean of 100,000 control networks ± s.d. Statistical significance is the fraction of control networks that exhibited the same or higher fraction of pairs with a high GO similarity or pairs present in WormNet as the actual interaction network. (**c**) Average Pearson correlation coefficient (PCC) score of mRNAs corresponding to protein pairs in the interaction data set (red arrow), compared with the distribution of average PCC scores of 100,000 control networks (blue line). Only AD-cDNA-derived pairs were analysed, as protein pairs in the AD-Fragment library already have very similar expression profiles (Supplementary Fig. 1d). Statistical significance is the fraction of control networks that exhibited an average PCC score identical to or higher than the actual interaction network. (**d**) Experimental validation by co-affinity purification. Bait proteins were expressed as Avi–mCherry fusions together with the bacterial biotin ligase BirA, which biotinylates the Avi sequence. Prey proteins were expressed as EGFP fusions. Bait proteins were purified from cell lysates using streptavidin-coated beads. Protein pairs tested are indicated above the blots. For prey proteins, f.l. indicates full-length protein, and frag. a shorter fragment. Lanes 1 and 2 are controls for nonspecific binding of bait to EGFP or prey to Avi–mCherry. The band with a relative molecular mass of ~35,000 (M_r 35K) in lane 2 is due to cross-reactivity of the anti-GFP antibody with Avi–mCherry. Asterisks indicate bands of expected molecular mass. Expression of tagged protein in input lysates is shown in Supplementary Fig. 2. Purifications were performed once. Unprocessed scans are shown in Supplementary Fig. 5. (**e**) Fraction of interactions from indicated data sets that test positive by MAPPIT at increasing assay stringency. *Ce*PRS, *Ce*RRS and WI-2007 data are reproduced from ref. 18. Shading indicates standard error of the proportion. At a scoring threshold of 12, $P < 1 \times 10^{-5}$ for each interaction data set compared with *Ce*RRS, two-sided Fisher's exact test.

C. elegans, or are homologous to known polarity regulators in other organisms (Fig. 1a and Supplementary Table 1). For each protein, a Y2H bait construct was generated by cloning the full-length open reading frame (ORF) into a Gal4 AD vector. We previously showed that a greater number of interactions can be detected when multiple fragments of a protein are tested in the Y2H system[9]. Therefore, up to 12 additional fragment bait constructs were cloned for each protein (Fig. 1b and Supplementary Table 1). We successfully generated 65 full-length bait clones, and 338 fragment bait clones, which together represent all 69 polarity regulators (Supplementary Table 1).

Each of the bait constructs was used to screen two Gal4 AD libraries: a library that contains full-length and fragment clones of 749 genes that are essential for early embryonic development (AD-Fragment library)[9], and a mixed-stage *C. elegans* AD-cDNA library (Fig. 1c). We eliminated auto-activators that arose during the screening process[10,11], and interactions where the AD-ORF fusion was out-of-frame. To further increase the accuracy, we included only AD-Fragment library-derived interactions identified in 2 or more yeast colonies. The AD-cDNA library is more complex, and many valid interactions may be identified only in a single yeast colony. Hence, we experimentally retested all interactions identified only once, retaining those that retested positively (Fig. 1c). The final *C. elegans* polarity interaction network (*Ce*PIN) contains 439 interactions between 296 proteins (Fig. 1d and Supplementary Table 2). Most interactions (359/439) were detected only using non-full-length bait constructs, confirming that the use of protein fragments increases the detectability of interactions[9]. The network contains 54 interactions that have previously been reported, including 19 interactions that have been studied in detail (Supplementary Table 3).

Quality assessment and experimental validation of the polarity interaction network

To assess the quality of the *Ce*PIN we examined whether interacting protein pairs shared other characteristics that indicate a functional

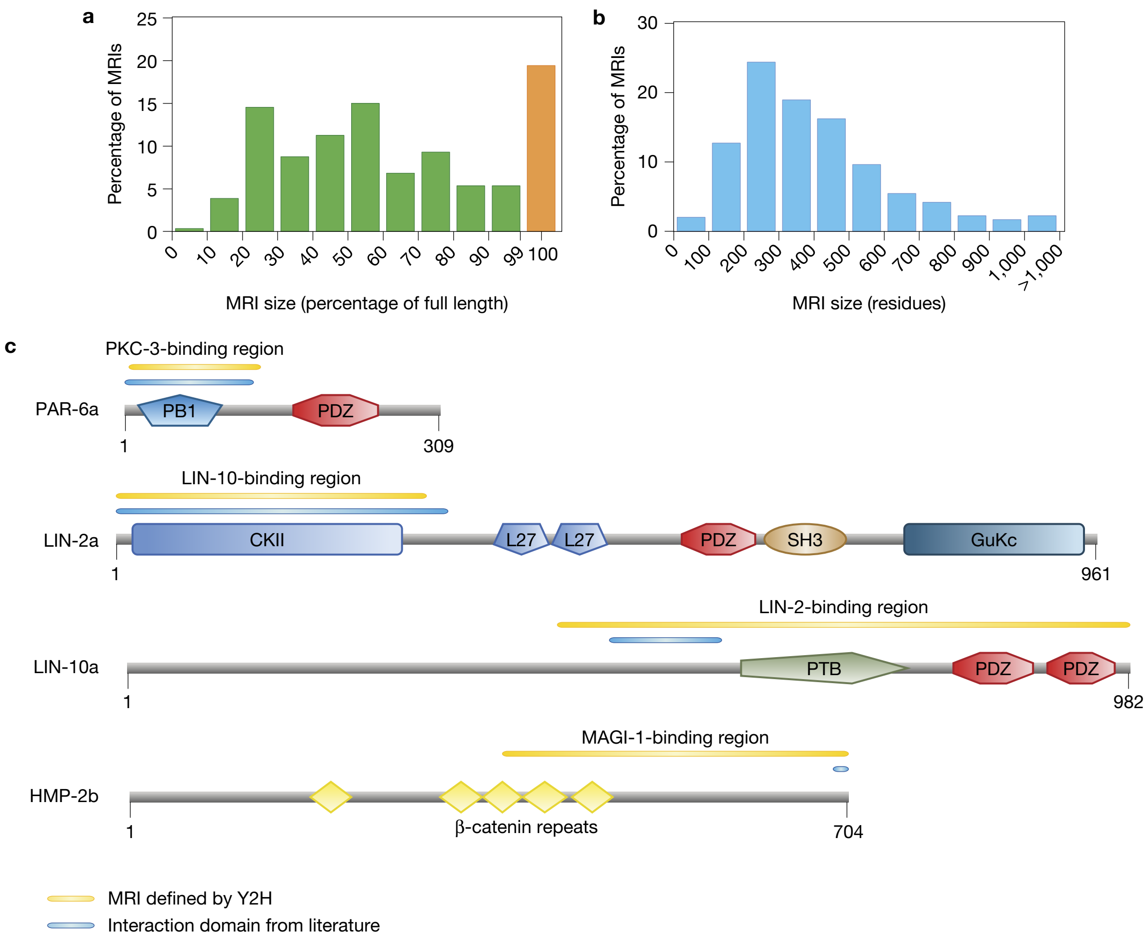


Figure 3 Identification and validation of minimal regions of interaction (MRIs). (**a**) Distribution of the size of the identified MRIs as a percentage of the full-length protein. Interactions identified only as full length are indicated separately (orange bar). (**b**) Distribution of the length of the identified MRIs in absolute residues. See Supplementary Fig. 6a–f for separate size distributions of MRIs identified from bait proteins, from AD-Fragment clones and from AD-cDNA clones. (**c**) Examples of MRIs for interactions where the binding site had already been described in the literature. See Supplementary Fig. 6g for the complete set of MRIs for interactions with known interaction sites.

association. We found that interacting protein pairs were enriched for similar GO terms, as well as for presence in WormNet, which predicts functional linkages between *C. elegans* genes[12] (Fig. 2a,b and Supplementary Fig. 1). In addition, interactions identified from AD-cDNA library screens were highly enriched for similar messenger RNA expression profiles (Fig. 2c and Supplementary Fig. 1d).

We next retested a random sample of 33 interactions by co-affinity purification from human embryonic kidney (HEK) 293 cells. Each interaction was tested using both a full-length prey protein and a shorter prey protein fragment containing the minimal region of interaction. We successfully reproduced the interaction between 16 of the 33 pairs tested (48%; Fig. 2d and Supplementary Fig. 2 and Supplementary Table 4). This is comparable to published validation rates of binary interactions in orthogonal assays[13,14] and to our own previous observations using this approach[15]. For 12 of the validated interactions, both the full-length and fragment prey were expressed. Of these, 6 were detected only using the fragment prey protein. Thus, the use of fragments provides an advantage in co-affinity purifications as well.

Finally, we retested 93 randomly selected protein pairs using the mammalian protein–protein interaction trap (MAPPIT), which detects interactions between proteins in living cells on the basis of restoration of the signalling capacity of a signalling-deficient cytokine receptor[16,17] (Supplementary Table 4). We examined the fraction that tested positively at a range of assay stringency thresholds, and compared the results with three previously published *C. elegans* MAPPIT analyses[18]: a positive reference set of 46 low-throughput literature-curated interactions (*Ce*PRS), a set of 86 random protein pairs that serves as a negative control (*Ce*RRS), and a set of 87 interactions detected in a large-scale Y2H screen (WI-2007; ref. 18). The retest rates of the *Ce*PIN, *Ce*PRS and WI-2007 were statistically indistinguishable (Fisher's exact tests), but scored significantly higher than the *Ce*RRS control pairs (Fig. 2e). Together with the identification of previously described interactions, the computational and experimental validations confirm the quality of the polarity interaction network.

Identification of minimal regions of interaction

For each interaction, we defined the minimal region of interaction (MRI) as the smallest region shared by all interacting protein fragments (Supplementary Table 5 and Supplementary Fig. 3). The

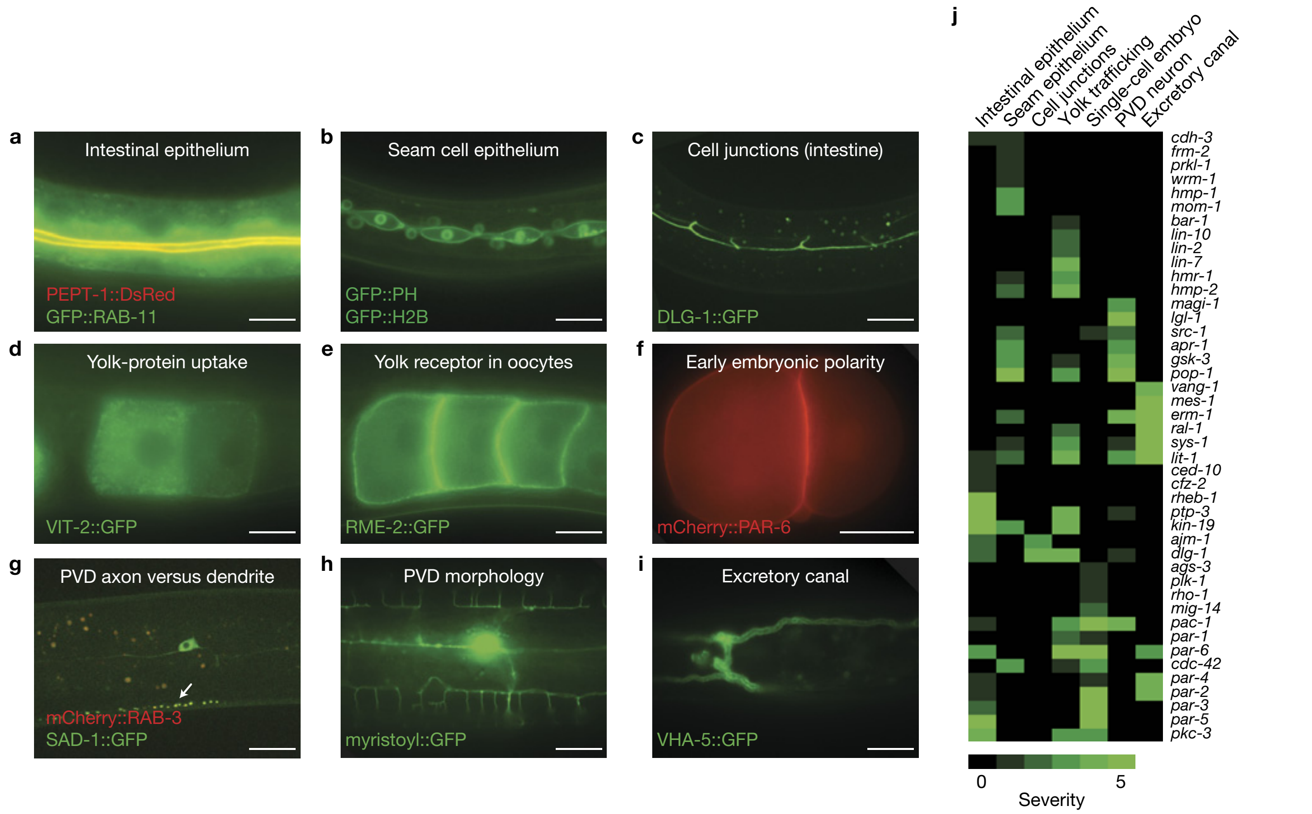


Figure 4 Phenotypic analysis of the bait proteins. (**a**–**i**) Marker lines used for phenotypic profiling. Scale bars, 10 μm. (**a**) Intestinal expression of PEPT-1::dsRed (apical domain) and GFP::RAB-11 (apical recycling endosomes)[25]. (**b**) Seam epithelium with cell outline marked by GFP fused to a pleckstrin homology domain, and nuclei marked by a histone 2B GFP fusion. (**c**) Expression of the junctional protein DLG-1::GFP (shown here in intestine). (**d**) Uptake of the yolk protein VIT-2::GFP in oocytes[26]. (**e**) Cortical localization of RME-2::GFP, the receptor for VIT-2, in oocytes. (**f**) Asymmetric distribution of mCherry::PAR-6 at the two-cell stage. (**g**) Localization of the presynaptic molecules mCherry::RAB-3 and SAD-1::GFP in the PVD axon (arrow). SAD-1::GFP is also observed at low levels in the cell body and dendrites near the cell body. (**h**) Expression of myristoylated GFP under the control of a PVD-specific promoter, to examine the extensive arborization pattern of the PVD dendritic tree. (**i**) Excretory canal outline by expression of VHA-5::GFP. Shown is the head region where the excretory canal cell body is located. (**j**) Hierarchical clustering of bait proteins on the basis of observed phenotypes. Green colour indicates the presence of a defect, and colour intensity indicates severity of the defect on a scale of 1–5. Phenotype scores were based on examination of >15 animals.

average length of all MRIs identified was 408 amino acids or 60% of the respective full-length protein (Fig. 3a,b and Supplementary Fig. 4). To evaluate the accuracy of the MRIs, we compared them with previously described interaction domains (Fig. 3c and Supplementary Fig. 4g and Supplementary Table 5). Out of 29 MRIs for which corresponding interaction domains were previously identified, 27 are consistent with the domain described (93%). In the two cases where the MRI differed from the described domain, the published interaction involved a mammalian homologue of the *C. elegans* protein. These may reflect a difference between the mammalian and *C. elegans* proteins. Further evidence of the accuracy of the MRIs comes from the co-affinity purification experiments, where 10/19 interactions for which the MRI was tested scored positively. Thus, the MRIs identified by Y2H were able to mediate the interaction in an orthogonal binary interaction assay as well.

Although some MRIs are a near exact match to the known interaction site, others span a larger protein region (Fig. 3c). One explanation is that shorter clones were not identified or are not present in the library. For example, the LIN-10 MRI that binds LIN-2 was defined from AD-cDNA clones, which can define only the amino-terminal MRI boundary. Alternatively, the interaction may be mediated by a short, linear motif that needs to be presented as part of a larger, folded polypeptide. For example, the HMP-2 MRI covered a much larger region than the carboxy-terminal four residues known to mediate the interaction with MAGI-1 (ref. 19). Especially in these cases, the MRIs identified by Y2H can provide a starting point for experiments requiring expression of an interacting fragment.

Phenotypic profiling of the polarity interaction network

Further evidence of a functional relationship can be obtained by integrating protein interaction data with phenotypic data[20–23]. We therefore performed phenotypic profiling by RNAi. Polarity regulators may act only in certain tissues or play highly distinct roles in polarity establishment. Therefore, we examined the effects of RNAi in nine different strains expressing fluorescently tagged proteins involved in several polarity-related processes (Supplementary Table 6 and Fig. 4a–i). To examine defects in early embryonic polarity, we used a strain expressing mCherry::PAR-6, which localizes anteriorly in the one-cell embryo and to the contact-free surfaces in young embryos. In the larval stages, we examined three different epithelial

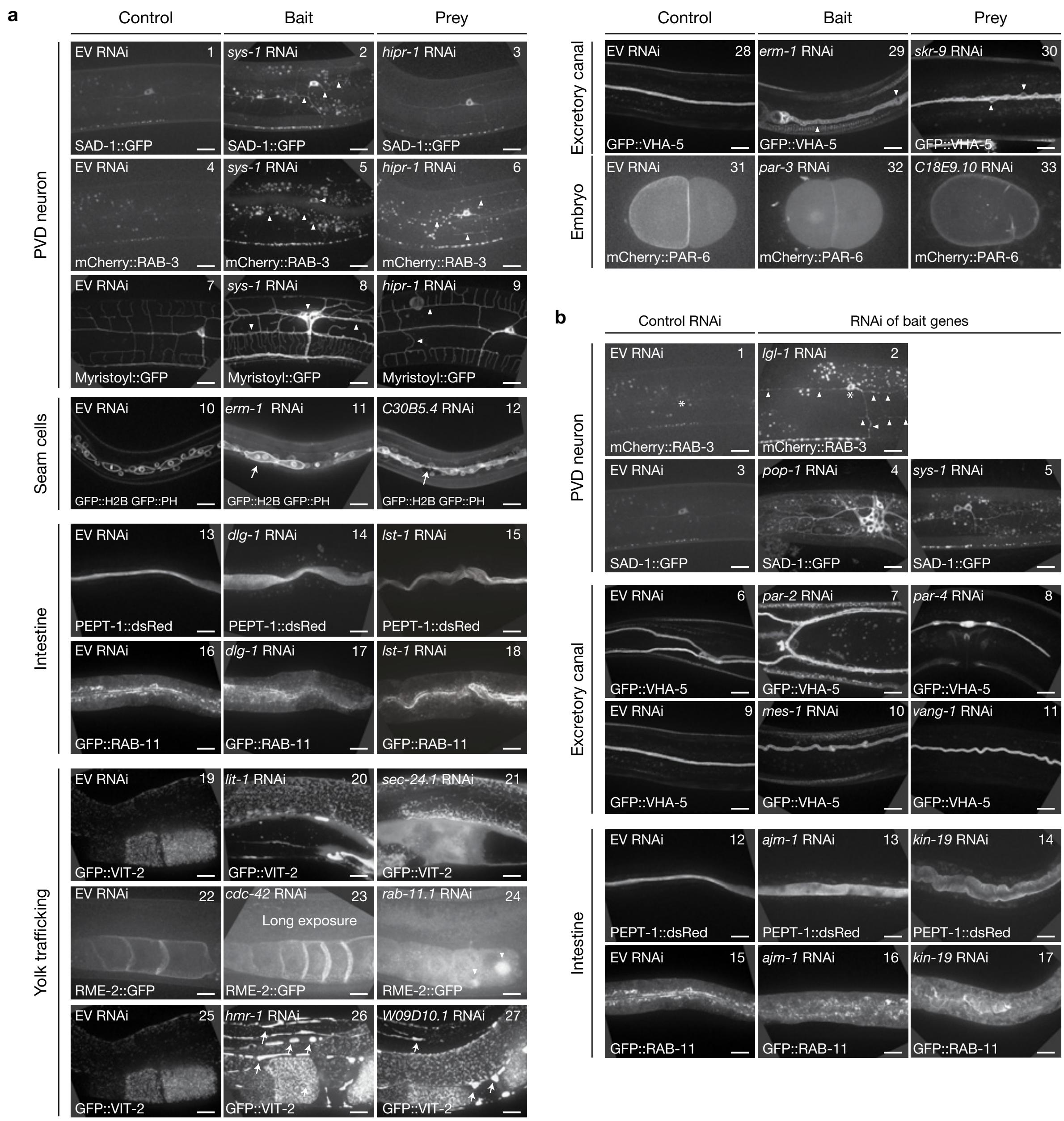


Figure 5 Examples of RNAi phenotypes. In all panels, the gene inactivated and the marker protein shown are indicated. EV: empty vector control RNAi. (**a**) Examples of interacting protein pairs where RNAi for both corresponding genes resulted in a defect in the same marker strain. In panel 2, arrowheads point to spreading of SAD-1::GFP into dendrites. In panels 4 and 5, arrowheads point to spreading of mCherry::RAB-3 into dendrites or cell body. In panels 8 and 9, arrowheads indicate defects that arise during the development of the PVD neuron. The arrow in panel 11 indicates failed cell division. The arrow in panel 12 points to a larger-than-normal distance between cells and elongated connection between cells. In the yolk-trafficking panels, arrows point to accumulation of yolk in the body cavity. In the excretory canal panels, arrowheads point to defects in canal morphology. (**b**) Examples of phenotypes observed in the PVD neuron, the excretory canal and the intestine during phenotypic analysis of the bait proteins. In PVD neuron panels, an asterisk indicates the location of the cell body, and arrowheads indicate spreading of RAB-3 into dendrites. Images are representative of phenotypes observed in >5 animals. Scale bars, 10 μm.

cell types: seam cells, intestinal cells, and the excretory canal. The seam cells exhibit both apical–basal polarity and anterior–posterior tissue polarity, and undergo a series of asymmetric stem cell-like divisions during larval development. The seam cell epithelium was examined in a strain expressing a GFP::PH domain fusion that outlines the cell membrane, and a GFP::H2B fusion that marks the DNA (ref. 24). Intestinal polarity was followed using a strain expressing PEPT-1::DsRed, which marks the apical domain, and GFP::RAB-11, which marks apically enriched recycling endosomes[25]. In both of these tissues we also examined the integrity of cell junctions using a DLG-1::GFP-expressing strain. The excretory canal is a tubular epithelium formed by a single cell, and therefore highly distinct from the multicellular seam and intestinal epithelia. We examined defects in excretory canal development using a strain expressing VHA-5::GFP.

In addition to these epithelial tissues, we examined two processes that depend on cell polarity. Uptake of yolk proteins by oocytes is an endocytic process that requires the activity of the *par-3*, *par-6*, *pkc-3* and *cdc-42* genes, and we examined this process in a strain expressing the yolk-protein fusion VIT-2::GFP (refs 26,27). All genes that showed a phenotype with this strain were re-screened in a strain expressing the fluorescently tagged VIT-2 receptor RME-2::GFP, to determine whether the observed defects were due to mislocalization of the receptor. Finally, neurons represent one of the most highly polarized cell types in the body. We examined the correct polarization of the axon in PVD neurons, in a strain expressing the fluorescently tagged presynaptic molecules mCherry::RAB-3 and SAD-1::GFP (ref. 28). All genes that showed a defect in localization of these markers following RNAi were re-screened in a strain expressing myristoylated GFP in the PVD neuron, to examine changes to the extensive arborization pattern of the PVD dendritic tree.

We first screened each of the 69 bait proteins for a total of 40 possible defects across the nine marker strains (Supplementary Table 6). For 44 bait proteins, downregulation by RNAi resulted in a detectable defect in at least one of the marker strains (Fig. 4j and Supplementary Table 6). Certain genes, for example, *par-6, kin-19, lit-1* and *dlg-1*, showed defects in most tissues examined, whereas others, such as *lgl-1* or *rheb-1*, affected only a single tissue. This matches our expectation that many genes will play tissue-specific roles in polarity establishment, although it may also reflect incomplete inactivation by RNAi. Next, the binding partners of these 44 bait proteins were screened specifically in those strains in which a defect was detected. For 100 protein pairs, RNAi of the corresponding genes resulted in a defect in the same polarity-related process (Fig. 5a and Supplementary Table 6). We found no bias among these 100 pairs for interactions found using full-length or fragment baits.

To evaluate whether an overlap in phenotype predicts functional association, we first used hierarchical clustering to group the bait genes by phenotypic similarity (Fig. 4j). Several genes known to act together also clustered together, including the *par* genes with *pkc-3* and *cdc-42*, the *lin-2,7,10* genes, whose protein products form a complex, and the genes encoding the interacting proteins AJM-1 and DLG-1. Next, we determined whether an overlap in phenotype correlates with protein interaction, with GO similarity, and with prior description of an interaction. We found that the interaction network was significantly enriched for an overlap in phenotype compared with all possible non-interacting pairs of proteins in the network (2.3-fold, $P = 3.0 \times 10^{-15}$). To examine the correlation with GO term similarity, we used semantic similarity scores calculated with the HRSS algorithm[29]. We defined a score of ≤ 0.1 as low similarity and >0.9 as high similarity. When examining all possible pairs of proteins in the network, we found that pairs that overlap in phenotype are enriched for a high GO similarity score (1.6-fold, $P = 0.001$) and are depleted of a low GO similarity score (1.8-fold, $P = 3.0 \times 10^{-20}$). When analysing only protein pairs that physically interact, we find that these pairs are already highly enriched for a high GO similarity score, which was not further enriched in the subset with an overlap in phenotype. However, this subset did show a significant depletion of pairs with low GO similarity (1.8-fold, $P = 0.006$). Thus, an overlap in phenotype positively correlates with high GO similarity, and negatively correlates with low GO similarity. Finally, we examined whether interacting proteins that overlap in phenotype were more likely to have been described in the literature. We found that 9% (9/100) of interactions with an overlap in phenotype had previously been described, compared with 2% (5/232) of the remaining interactions (Supplementary Table 6). Taken together, these data indicate that the phenotypic profiling is able to capture interactions relevant *in vivo* and to cell polarity.

A strength of unbiased interaction mapping is that it may produce leads into the function of proteins. Indeed, many of the interactions with an overlap in phenotype involve proteins not known to act in a common process. For example, we identified an interaction between SYS-1, a *C. elegans* β-catenin homologue, and HIPR-1, the *C. elegans* homologue of mammalian Huntingtin-interacting protein 1 (HIP1) and HIP1-related (HIP1R), which are thought to function in the endocytic pathway[30]. Inactivation of either *sys-1* or *hipr-1* resulted in spreading of SAD-1 and RAB-3 vesicles into the dendrites of the PVD neuron (Fig. 5a panels 1–6). A neuronal defect is consistent with previous observations that *hipr-1* modulates the presynaptic activity of *C. elegans* neurons, and with reported localization of HIP1 and HIPR1 to dendritic structures[30,31]. However, SYS-1 has only been described to function as a transcriptional co-activator with POP-1 TCF (refs 32,33). Interestingly, SYS-1 also interacted with UNC-11, a homologue of the AP180 clathrin adaptor, and RNAi against both *sys-1* and *unc-11* resulted in defects in yolk endocytosis (Supplementary Table 6). Not all phenotypes overlapped closely. For example, we identified an interaction between the Ezrin–Radixin–Moesin homologue ERM-1 and the uncharacterized protein C30B5.4. RNAi for both genes resulted in defects in the seam epithelium, although the nature of the phenotype differed (Fig. 5a panels 10–12). C30B5.4 contains an RNA-binding motif, and is homologous to human RBMX2, which is annotated as a spliceosome component. As a final example, we identified an interaction between DLG-1 Discs Large and LST-1, a nematode-specific protein. RNAi for both *dlg-1* and *lst-1* resulted in irregular morphology of the intestinal lumen (Fig. 5a panels 13–15), and a reduction in the subapical accumulation of RAB-11 vesicles (Fig. 5a panels 16–18). Consistent with an intestinal defect, LST-1 is expressed in the intestine[34]. For each of these interactions, the physical association and overlap in phenotype suggest a functional association, even though it is not immediately obvious what the functional consequences of this association might be.

The phenotypic profiling of the 69 bait proteins led to the identification of several phenotypes not previously described. For example, inactivation of *lgl-1*, the *C. elegans* homologue of *Drosophila lethal giant larvae*, caused spreading of RAB-3 vesicles into the dendrites of the PVD neuron (Fig. 5b panel 2). *C. elegans lgl-1* was shown to act redundantly with *par-2* in early embryonic polarity establishment[35,36], but a *par-2*-independent function for *lgl-1* had not yet been described. A pronounced effect on the PVD neuron was observed on inactivation of the TCF transcription factor *pop-1,* which resulted in the formation of numerous additional PVD cell bodies and spreading of RAB-3 vesicles into dendrites (Fig. 5b panel 4). This may reflect additional cell divisions or cell-fate alterations in the V5 lineage that produces the PVD neuron, rather than a polarity defect, as *pop-1* is known to be required for cell-fate decisions. Inactivation of SYS-1, which binds POP-1, also resulted in spreading of RAB-3 and duplication of the cell body (Fig. 5b panel 5).

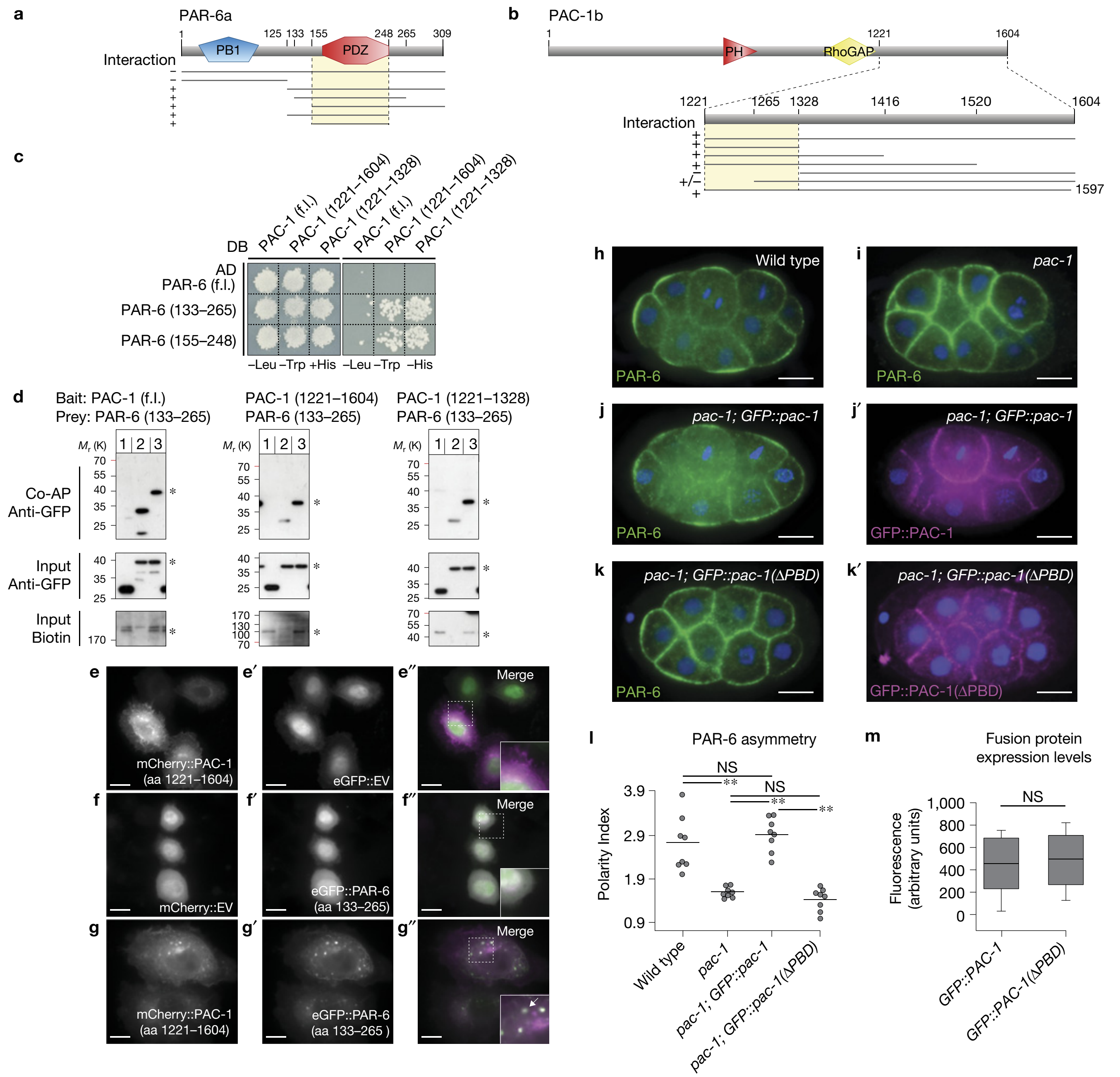


Figure 6 The PAR-6–PAC-1 interaction is required for radial polarization. (**a**–**c**) Delineation of the protein regions that mediate the interaction between PAR-6 and PAC-1. (**a**,**b**) Schematic representations of protein fragments (grey lines) tested for interaction by Y2H. Presence (+) or absence (−) of interaction is indicated to the left of each fragment. Yellow regions represent the minimal regions of interaction. (**c**) Example of Y2H analysis using the fragments indicated. Left: permissive plate containing histidine; right: selective plate lacking histidine. Representative image of 3 experiments. (**d**) Validation of the PAC-1–PAR-6 interaction and the interaction domains by co-purification. Western blots show co-purification of GFP::PAR-6 with biotinylated mCherry::PAC-1. Protein fragments tested are indicated. See legend of Fig. 2d for details. Co-purifications were performed at least twice. Unprocessed scans are shown in Supplementary Fig. 5. (**e**–**g**) Subcellular localization of mCherry::PAC-1 (amino acids 1221–1604) and EGFP::PAR-6 (amino acids 133–155) in HeLa cells. Insets are magnified ×2. Localization patterns were observed in >30 cells. (**h**,**i**) PAR-6 is enriched at contact-free surfaces in wild-type embryos ($n=58/58$) and localizes symmetrically to cell surfaces in *pac-1* mutant embryos ($n=62/62$). (**j**,**j'**) PAR-6 is enriched at contact-free surfaces in *pac-1; GFP::pac-1* embryos ($n=50/50$). (**j'**) GFP::PAC-1 localizes to sites of cell contact ($n=50/50$). (**k**) PAR-6 localizes symmetrically to cell surfaces in *pac-1; GFP::pac-1(ΔPBD)* embryos ($n=61/61$). (**k'**) GFP::PAC-1(ΔPBD) localizes to sites of cell contact, similarly to GFP::PAC-1 ($n=61/61$). (**l**) Quantification of PAR-6 asymmetry in embryos of the indicated genotypes (see Methods for details). Circles represent individual embryos and the bar indicates mean value. Eight embryos were measured for each genotype. *P* values were calculated using the Mann–Whitney *U*-test. NS, not significantly different; $^{**}P < 0.01$. Samples were pooled from three independent experiments. (**m**) Expression levels of GFP::PAC-1 and GFP::PAC-1(ΔPBD) measured as fluorescence intensity at the four-cell stage. Statistical analysis was performed using the Mann–Whitney *U*-test, $n=7$ GFP::PAC-1 embryos and $n=9$ GFP::PAC-1(ΔPBD) embryos. Box represents first and third quartiles, bars represent maximum and minimum values, and the line within the box is the mean fluorescence intensity. Samples were pooled from two independent experiments. Scale bars, 10 μm.

We identified nine genes whose inactivation resulted in defects in the excretory canal. Three of these, *erm-1*, *par-6* and *ral-1*, had previously described roles in excretory canal formation[37–39]. We also identified defects in canal appearance for *par-2* and *par-4*, which indicates a broader involvement of cell polarity regulators in excretory canal formation (Fig. 5 panels 7 and 8). Finally, *mes-1* and *vang-1* showed a defect exclusively in the excretory canal (Fig. 5b panels 10 and 11). MES-1 is a receptor tyrosine kinase-like protein required for accurate positioning of the mitotic spindle in the P2 and EMS cells of the early embryo[40,41]. VANG-1 is the *C. elegans* homologue of the planar cell polarity regulator Strabismus (also known as Van Gogh)[42–44]. The identification of roles in excretory canal formation for *mes-1* and *vang-1* may provide insights into the formation of this organ.

Interaction between PAR-6 and PAC-1 is required for radial polarization

As a final demonstration of the validity of our approach to identify functionally relevant interactions, we focused on the interaction between PAR-6 and PAC-1. In addition to the physical interaction, the phenotypic profiles of *par-6(RNAi)* and *pac-1(RNAi)* clustered together (Fig. 4j). PAC-1 ARHGAP21 is a RhoGAP protein required for radial polarization of the *C. elegans* embryo[45]. In wild-type embryos, PAR-6 localizes to the outer, contact-free cell surface beginning at the late 4-cell stage. Cortical recruitment of PAR-6 is dependent on the Rho-family member CDC-42, which localizes along the entire cortex[45,46]. PAC-1 localizes to cell contact sites, and is thought to trigger radial polarization by locally inactivating CDC-42, limiting the recruitment of PAR-6 by CDC-42 to the outer cell surfaces[45,46]. Although PAR-6 and PAC-1 do not seem to co-localize in fully polarized cells, they are both found at cell contacts as polarity is initially established, and our observations suggest that a physical interaction between PAR-6 and PAC-1 contributes to their functions.

Additional fragments of PAR-6 and PAC-1 were tested by Y2H to narrow down the interaction sites. The PDZ domain of PAR-6 (amino acids 155–248) was sufficient to mediate binding to PAC-1 (Fig. 6a,c). A PAC-1 fragment lacking the final 7 residues was still able to interact with PAR-6, indicating that the PAR-6 PDZ domain interacts with an internal motif of PAC-1 (Fig. 6b,c). A 100-amino-acid sequence just downstream of the PAC-1 RhoGAP domain (amino acids 1221–1328) was able to mediate binding to PAR-6 (Fig. 6b,c). We confirmed the interaction between PAR-6 and PAC-1 by co-affinity purification from mammalian HEK293 cells. A C-terminal fragment of PAC-1b (amino acids 1221–1604), as well as the minimal PAR-6-binding domain (amino acids 1221–1328), co-purified with a PAR-6a fragment containing the PDZ domain (amino acids 133–265; Fig. 6d). Moreover, when co-expressed in HeLa cells, PAC-1 and PAR-6 co-localize in a punctate pattern that was clearly distinct from the localization pattern of PAC-1 or PAR-6 expressed individually (Fig. 6e–g). These observations strongly support the presence of a physical interaction between PAC-1 and PAR-6.

To explore the functional relevance of the interaction between PAC-1 and PAR-6, we determined whether PAC-1 lacking the PAR-6-binding domain can functionally substitute for the wild-type protein. As previously demonstrated[45], expression of a GFP::PAC-1 fusion protein rescues the localization of PAR-6 in a *pac-1* mutant embryo (Fig. 6j). We deleted the PAR-6-binding domain (PBD; amino acids 1221–1328) from the GFP::PAC-1-encoding plasmid and expressed GFP::PAC-1(ΔPBD) in *pac-1(xn6)* mutant embryos. GFP::PAC-1(ΔPBD) was expressed at similar levels to GFP::PAC-1 (Fig. 6m) and localized to the inner cell surfaces (Fig. 6k′). However, GFP::PAC-1(ΔPBD) failed to restore the radial localization pattern of PAR-6 (Fig. 6k,l). Thus, the direct interaction of PAC-1 with PAR-6 is necessary for PAC-1 to establish radial polarization.

DISCUSSION

In this study we identified a *C. elegans* polarity interaction network of 439 interactions between 296 proteins, and validated the quality of the network using a combination of computational and experimental approaches. The experimental identification of the protein regions that mediated these interactions provides additional insight into the proteins in the network, as well as a starting point for studies needing to express functional protein domains.

Of the 439 interactions we identified, 385 had not been previously identified. Furthermore, out of 19 interactions studied in detail, only 12 had previously been identified in large-scale studies. These results confirm the ability of a smaller-scale targeted approach to identify additional relevant interactions not present in high-throughput interaction data sets.

One hundred protein pairs physically interacted and, on RNAi-mediated depletion, showed a defect in the same polarity-related process. An overlap in phenotype correlated with the presence of a physical interaction, with GO similarity, and with prior description of an interaction in the literature, confirming that the combination of interaction data with phenotypic data better predicts a functional relationship than presence of an interaction alone. The marker strains we used for the phenotypic profiling highlight only a subset of polarity-related processes. Thus, examining the interacting protein pairs using other marker lines should result in the identification of additional pairs with overlapping phenotypes.

The *Ce*PIN provides many putative polarity-related interactions, which can be pursued using detailed *in vivo* analyses that can identify more subtle or tissue-specific roles in cell polarity. Many of the mechanisms that drive polarity establishment are conserved between *C. elegans*, flies and humans, and interactions between conserved proteins can be examined in other model systems as well. Thus, the *Ce*PIN will be a valuable resource for future studies aiming to gain new insights into the mechanisms that control cell polarity. □

METHODS

Methods and any associated references are available in the online version of the paper.

Note: Supplementary Information is available in the online version of the paper

ACKNOWLEDGEMENTS

We thank M. Zerial for providing the MZE1 strain, C. Bargmann for providing strain CX9797 and a strain carrying the *Pdes-2::myristoyl::GFP* extrachromosomal array, and M. Harterink for strain STR62. We thank S. de Rouck for technical assistance, and M. Vidal for permission to reanalyse MAPPIT data. We thank F. Zwartkruis, M. Gloerich and A. Thomas for critically reading the manuscript. Some nematode strains used in this work were provided by the Caenorhabditis Genetics Center, which is funded by the NIH National Center for Research Resources (NCRR). This work was supported by the Netherlands Organization for Scientific Research (NWO) ALW Innovational Research Incentives Scheme Vidi grant 864.09.008 to M.B., NWO-CW ECHO grant 711.014.005 to M.B., National Science Foundation Graduate Research Fellowship 12-A0-00-000165-01 to D.K., and NIH grants R01GM078341 and R01GM098492 to J.N. J.T. is the recipient of ERC Advanced Grant 340941.

AUTHOR CONTRIBUTIONS

M.v.d.V. contributed to the cloning and Y2H screens. D.K. performed the PAC-1 rescue experiments. I.L. performed the MAPPIT experiments. J.J.R. and S.N. contributed to the RNAi screens. M.B. performed the computational analyses. All other experiments were performed by T.K. S.v.d.H. and T.K. contributed to the design of the study. J.N. contributed to the design of the PAC-1 PAR-6 experiments. J.T. contributed to the design of the MAPPIT experiments. T.K. and M.B. wrote the manuscript. M.B. guided all aspects of the study.

COMPETING FINANCIAL INTERESTS

The authors declare no competing financial interests.

Published online at http://dx.doi.org/10.1038/ncb3300
Reprints and permissions information is available online at www.nature.com/reprints

1. Humbert, P. O. *et al.* Control of tumourigenesis by the Scribble/Dlg/Lgl polarity module. *Oncogene* **27,** 6888–6907 (2008).
2. Tepass, U. The apical polarity protein network in *Drosophila* epithelial cells: regulation of polarity, junctions, morphogenesis, cell growth, and survival. *Annu. Rev. Cell Dev. Biol.* **28,** 655–685 (2012).
3. St Johnston, D. & Ahringer, J. Cell polarity in eggs and epithelia: parallels and diversity. *Cell* **141,** 757–774 (2010).
4. Jacob, L., Opper, M., Metzroth, B., Phannavong, B. & Mechler, B. M. Structure of the *l(2)gl* gene of *Drosophila* and delimitation of its tumor suppressor domain. *Cell* **50,** 215–225 (1987).
5. Woods, D. F. & Bryant, P. J. The *discs-large* tumor suppressor gene of *Drosophila* encodes a guanylate kinase homolog localized at septate junctions. *Cell* **66,** 451–464 (1991).
6. Bilder, D. & Perrimon, N. Localization of apical epithelial determinants by the basolateral PDZ protein Scribble. *Nature* **403,** 676–680 (2000).
7. Goehring, N. W. & Grill, S. W. Cell polarity: mechanochemical patterning. *Trends Cell Biol.* **23,** 72–80 (2013).
8. Rodriguez-Boulan, E. & Macara, I. G. Organization and execution of the epithelial polarity programme. *Nat. Rev. Mol. Cell Biol.* **15,** 225–242 (2014).
9. Boxem, M. *et al.* A protein domain-based interactome network for *C. elegans* early embryogenesis. *Cell* **134,** 534–545 (2008).
10. Vidalain, P.-O., Boxem, M., Ge, H., Li, S. & Vidal, M. Increasing specificity in high-throughput yeast two-hybrid experiments. *Methods* **32,** 363–370 (2004).
11. Walhout, A. J. & Vidal, M. A genetic strategy to eliminate self-activator baits prior to high-throughput yeast two-hybrid screens. *Genome Res.* **9,** 1128–1134 (1999).
12. Cho, A. *et al.* WormNet v3: a network-assisted hypothesis-generating server for Caenorhabditis elegans. *Nucleic Acids Res.* **42,** W76–W82 (2014).
13. Braun, P. *et al.* An experimentally derived confidence score for binary protein–protein interactions. *Nat. Methods* **6,** 91–97 (2009).
14. Rolland, T. *et al.* A proteome-scale map of the human interactome network. *Cell* **159,** 1212–1226 (2014).
15. Waaijers, S., Koorman, T., Kerver, J. & Boxem, M. Identification of human protein interaction domains using an ORFeome-based yeast two-hybrid fragment library. *J. Proteome Res.* **12,** 3181–3192 (2013).
16. Eyckerman, S. *et al.* Design and application of a cytokine-receptor-based interaction trap. *Nat. Cell Biol.* **3,** 1114–1119 (2001).
17. Lemmens, I., Lievens, S. & Tavernier, J. MAPPIT, a mammalian two-hybrid method for in-cell detection of protein–protein interactions. *Methods Mol. Biol.* **1278,** 447–455 (2015).
18. Simonis, N. *et al.* Empirically controlled mapping of the *Caenorhabditis elegans* protein–protein interactome network. *Nat. Methods* **6,** 47–54 (2009).
19. Stetak, A., Hörndli, F., Maricq, A. V., van den Heuvel, S. & Hajnal, A. Neuron-specific regulation of associative learning and memory by MAGI-1 in *C. elegans*. *PLoS ONE* **4,** e6019 (2009).
20. Gunsalus, K. C. *et al.* Predictive models of molecular machines involved in Caenorhabditis elegans early embryogenesis. *Nature* **436,** 861–865 (2005).
21. Boulton, S. J. *et al.* Combined functional genomic maps of the *C. elegans* DNA damage response. *Science* **295,** 127–131 (2002).
22. Tu, Z. *et al.* Integrating siRNA and protein–protein interaction data to identify an expanded insulin signaling network. *Genome Res.* **19,** 1057–1067 (2009).
23. Wang, L., Tu, Z. & Sun, F. A network-based integrative approach to prioritize reliable hits from multiple genome-wide RNAi screens in *Drosophila*. *BMC Genomics* **10,** 220 (2009).
24. Wildwater, M., Sander, N., de Vreede, G. & van den Heuvel, S. Cell shape and Wnt signaling redundantly control the division axis of *C. elegans* epithelial stem cells. *Development* **138,** 4375–4385 (2011).
25. Winter, J. F. *et al. Caenorhabditis elegans* screen reveals role of PAR-5 in RAB-11-recycling endosome positioning and apicobasal cell polarity. *Nat. Cell Biol.* **14,** 666–676 (2012).
26. Grant, B. & Hirsh, D. Receptor-mediated endocytosis in the Caenorhabditis elegans oocyte. *Mol. Biol. Cell* **10,** 4311–4326 (1999).
27. Balklava, Z., Pant, S., Fares, H. & Grant, B. D. Genome-wide analysis identifies a general requirement for polarity proteins in endocytic traffic. *Nat. Cell Biol.* **9,** 1066–1073 (2007).
28. Maniar, T. A. *et al.* UNC-33 (CRMP) and ankyrin organize microtubules and localize kinesin to polarize axon-dendrite sorting. *Nat. Neurosci.* **15,** 48–56 (2012).
29. Wu, X., Pang, E., Lin, K. & Pei, Z.-M. Improving the measurement of semantic similarity between gene ontology terms and gene products: insights from an edge- and IC-based hybrid method. *PLoS ONE* **8,** e66745 (2013).
30. Gottfried, I., Ehrlich, M. & Ashery, U. The Sla2p/HIP1/HIP1R family: similar structure, similar function in endocytosis? *Biochem. Soc. Trans.* **38,** 187–191 (2010).
31. Parker, J. A. *et al.* Huntingtin-interacting protein 1 influences worm and mouse presynaptic function and protects Caenorhabditis elegans neurons against mutant polyglutamine toxicity. *J. Neurosci.* **27,** 11056–11064 (2007).
32. Liu, J. *et al.* The C. elegans SYS-1 protein is a bona fide β-catenin. *Dev. Cell* **14,** 751–761 (2008).
33. Phillips, B. T., Kidd, A. R., King, R., Hardin, J. & Kimble, J. Reciprocal asymmetry of SYS-1/beta-catenin and POP-1/TCF controls asymmetric divisions in Caenorhabditis elegans. *Proc. Natl Acad. Sci. USA* **104,** 3231–3236 (2007).
34. Cuppen, E., van der Linden, A. M., Jansen, G. & Plasterk, R. H. A. Proteins interacting with Caenorhabditis elegans Galpha subunits. *Comp. Funct. Genomics* **4,** 479–491 (2003).
35. Hoege, C. *et al.* LGL can partition the cortex of one-cell Caenorhabditis elegans embryos into two domains. *Curr. Biol.* **20,** 1296–1303 (2010).
36. Beatty, A., Morton, D., Kemphues, K. & The, C. elegans homolog of Drosophila Lethal giant larvae functions redundantly with PAR-2 to maintain polarity in the early embryo. *Development* **137,** 3995–4004 (2010).
37. Khan, L. A. *et al.* Intracellular lumen extension requires ERM-1-dependent apical membrane expansion and AQP-8-mediated flux. *Nat. Cell Biol.* **15,** 143–156 (2013).
38. Göbel, V., Barrett, P. L., Hall, D. H. & Fleming, J. T. Lumen morphogenesis in *C. elegans* requires the membrane-cytoskeleton linker erm-1. *Dev. Cell* **6,** 865–873 (2004).
39. Armenti, S. T., Chan, E. & Nance, J. Polarized exocyst-mediated vesicle fusion directs intracellular lumenogenesis within the *C. elegans* excretory cell. *Dev. Biol.* **394,** 110–121 (2014).
40. Berkowitz, L. A. & Strome, S. MES-1, a protein required for unequal divisions of the germline in early *C. elegans* embryos, resembles receptor tyrosine kinases and is localized to the boundary between the germline and gut cells. *Development* **127,** 4419–4431 (2000).
41. Bei, Y. *et al.* SRC-1 and Wnt signaling act together to specify endoderm and to control cleavage orientation in early *C. elegans* embryos. *Dev. Cell* **3,** 113–125 (2002).
42. Hoffmann, M., Segbert, C., Helbig, G. & Bossinger, O. Intestinal tube formation in Caenorhabditis elegans requires vang-1 and egl-15 signaling. *Dev. Biol.* **339,** 268–279 (2010).
43. Honnen, S. J. *et al.* C. elegans VANG-1 modulates life span via insulin/IGF-1-like signaling. *PLoS ONE* **7,** e32183 (2012).
44. Sanchez-Alvarez, L. *et al.* VANG-1 and PRKL-1 cooperate to negatively regulate neurite formation in Caenorhabditis elegans. *PLoS Genet.* **7,** e1002257 (2011).
45. Anderson, D. C., Gill, J. S., Cinalli, R. M. & Nance, J. Polarization of the *C. elegans* embryo by RhoGAP-mediated exclusion of PAR-6 from cell contacts. *Science* **320,** 1771–1774 (2008).
46. Chan, E. & Nance, J. Mechanisms of CDC-42 activation during contact-induced cell polarization. *J. Cell Sci.* **126,** 1692–1702 (2013).

METHODS

Primers and Linkers. All primers and oligonucleotide linkers are listed in Supplementary Table 7.

Yeast two-hybrid vectors. Vectors pMB28 and pMB29 were derived from pPC97 and pPC86 respectively[47,48]. Both encode a flexible linker (GGGG) and AscI or NotI restriction sites, inserted as an oligonucleotide linker (linker 1) into the parent vectors digested with SmaI and SacI. pMB28 also contains the *CAN1* gene, inserted into the ApaI site as a PCR product amplified from yeast genomic DNA (primers CAN1_F and CAN1_R). pMB29 contains the *CYH2* gene, cut from Clontech vector pAS2-1 with EcoRV and ligated into vector digested with Acc65I and treated with Klenow + dNTPs to generate blunt ends.

Generating DB-ORF bait clones. Full-length clones were amplified from cDNA with KOD polymerase (Novagen), using primers with AscI and NotI extension tails. PCR products were gel-purified, digested with AscI or NotI, and ligated into linearized pMB28. Clones were verified using internal gene-specific primers and primers DB and TERM. Fragment clones were designed on the basis of the size of the gene and the protein domain architecture as predicted by Pfam (http://pfam.sanger.ac.uk) and SMART (http://smart.embl-heidelberg.de). Fragments were PCR amplified using full-length sequence-verified clones as templates. For 7 genes for which we were unable to obtain a full-length clone, cDNA was used as the template. For each DB-fragment, six clones with a correct insert size were pooled before transformation into yeast.

Yeast two-hybrid analysis. DB-ORF (bait) clones were transformed into *Saccharomyces cerevisiae* strain Y8930 (MATα)[49] using the Te-LiAc method[50]. Bait strains able to activate reporter genes in the absence of AD-plasmid were discarded as auto-activators. Yeast two-hybrid (Y2H) screens were done as previously described[9]. Two Y2H AD libraries were screened: a library carrying fragments of 749 genes required for early embryogenesis[9], and a mixed-stage cDNA library (a gift from X. Xin and C. Boone, University of Toronto, Canada). To generate the AD mating libraries, yeast strain Y8800 (MAT**a**) was transformed with 30 μg of each library. After three days of growth at 30 °C, colonies were collected in YEPD + 20% glycerol, and frozen in 1 ml aliquots. Counter-selection on plates containing cycloheximide was used to eliminate auto-activators that arose during the screening process[9].

Identification of prey protein identity. Gal4 AD vector inserts were PCR amplified from yeast using primers AD and TERM as described previously[9], and DNA sequencing was performed by Macrogen Europe (http://dna.macrogen.com). Sequences were analysed by phred[51], using the default settings and the –trim_alt flag. Bases with quality scores below the default threshold of 0.5 were eliminated. Vector ends were clipped from the traces. Sequences where no 5′ vector end could be identified were eliminated. The identity of the ORF was determined by DNA BLAST analysis against a database containing all protein-coding genes in Wormbase release WS235. To determine whether the ORF was in-frame with the Gal4 AD sequence, we determined the identity of the ORF by protein BLAST using the translated protein sequence. A trace was considered in-frame if the DNA and protein BLAST identities match, or when no stop codon was encountered in the first 100 codons. Out-of-frame sequences were eliminated.

Verification of interactions. Protein pairs identified a single time from the AD-cDNA library were retested by isolating the Gal4 AD plasmid from the original yeast clone, transforming the isolated plasmid into yeast strain Y8800, and mating with the corresponding Y8930 bait strain. *HIS3* reporter activity was assayed on –Leu –Trp –His plates, and protein pairs that failed to activate the reporter were not included in the data set. For interactions identified multiple times, we confirmed the bait protein identity by PCR amplification and sequencing of the bait construct from a representative yeast colony.

PubMed searches. PubMed searches for interactions between *C. elegans* proteins were guided by automated searches of the Textpresso database (http://www.textpresso.org/nematode) for the co-occurrence within the same publication of two interacting proteins. Publications were then read to determine whether an interaction was described. We also manually searched PubMed for interactions between homologous proteins in other organisms.

Comparison with IntAct. To identify interactions present in IntAct (https://www.ebi.ac.uk/intact), we queried IntAct release 27. To identify orthologous interactions, we compiled a list of potential homologues from Homologene build 67 (http://www.ncbi.nlm.nih.gov/homologene), Inparanoid release 7 (http://inparanoid.sbc.su.se/cgi-bin/index.cgi), Orthomcl version 5 (http://orthomcl.org/orthomcl), and Ensembl release 73 (http://www.ensembl.org/index.html).

Control networks used for computational analysis. To determine whether the *Ce*PIN was enriched in similar GO terms, similar expression profiles, or presence in WormNet, we compared the interaction network with 100,000 control networks. We screened two very different AD libraries with correspondingly different search spaces: a genome-wide cDNA library and the early embryogenesis AD-Fragment library, which contains only 749 proteins that already have very similar expression profiles and gene ontology descriptions. AD-cDNA- and early embryogenesis AD-Fragment-derived interactions were therefore analysed separately. The control networks were generated by replacing the prey names of the interaction network with protein names randomly selected from the proteins present in the Gal4 AD library screened. This maintains the topology of the original network. Homodimers were excluded from all analyses.

GO term analysis. We used the Hybrid Relative Specificity Similarity based on Gene Ontology (HRSS) software[29] in combination with GO database release 201310 (http://www.geneontology.org) to calculate GO semantic similarity scores, limiting our analysis to the Biological Process domain. We compared the distribution of similarity scores for interacting protein pairs with the average distribution of scores of protein pairs in the corresponding control networks. Statistical significance is the fraction of the 100,000 control networks that exhibited the same fraction of pairs with a particular similarity score as the actual interaction network.

Comparison with WormNet. The WormNet v3 data set[12] was queried for the presence of gene pairs, and to extract the log-likelihood score (LLS) of a true functional linkage. Enrichment values are the fraction of interactions identified by Y2H that are present in WormNet divided by the average fraction of the control network pairs present in WormNet. Statistical significance was calculated as the fraction of the 100,000 control networks that exhibited the same or higher fraction of pairs present in WormNet as the actual interaction network.

Gene expression profiling comparison. We calculated pair-wise Pearson correlation coefficients between all gene pairs across all *C. elegans* microarray data collected in Wormbase release WS236. Data were downloaded from the SPELL website (http://spell.caltech.edu:3000). For each data set, we generated normalized Fisher z-transformed pair-wise correlation scores as described previously[52]. We excluded data sets with fewer than five different conditions, as well as data sets where >10% of genes had exactly the same expression levels in all conditions. The average of all z correlations was calculated, and converted back to PCC values. Gene pairs that had data in less than 50% of the data sets were excluded. The final compendium contains data for 17,287 protein-coding genes compiled from 110 data sets and 2,394 conditions. All calculations and plotting of PCC value distributions were done in the R language[53]. Statistical significance was calculated as the fraction of the 100,000 control networks that had the same or higher average PCC value as the actual interaction network.

Co-affinity purification. Co-affinity purifications were done as described previously[15]. As bait, the protein fragment that most frequently detected the interaction by Y2H was used. To clone ORFs into vectors Avi–mCherry–C1 and pEGFP–C1, sequences were PCR amplified from a cDNA using Hot Start KOD polymerase (Novagen) and primers with appropriate restriction enzyme tails. PCR products were gel-purified, digested, and cloned into linearized vectors. Enzyme combinations used were BglII+Acc65I, XhoI+HindIII, SalI+Acc65I, SalI+BamHI and SmaI+XmaI. HEK293 cells were originally obtained from the ATCC (http://www.lgcstandards-atcc.org). Their identity was validated by light microscopic observation. Cells were regularly tested for mycoplasma infection.

MAPPIT. Experiments were performed in HEK293T cells, which were authenticated in ref. 54, and regularly tested for mycoplasma infection. We generated Gateway entry clones of the bait and prey ORFs by PCR amplification and BP cloning into vector pDONR223. For both bait and prey constructs, we cloned the fragment that was found most frequently in the Y2H screens. The entry clones were transferred into the two MAPPIT vectors, and MAPPIT assays were performed as described previously[9]. The fraction of positives was scored over an experiment to control ratio (ECR) threshold range of 1 to 19.

Defining minimal regions of interaction. Minimal regions of interaction (MRIs) for each interaction were defined as the region of a protein that is present in all of the fragments found to interact. When multiple splice variants were predicted for a gene, the splice variant we identified most frequently was used. In cases where multiple fragments of a protein without any overlap were found to interact, no MRI was defined.

 DOI: 10.1038/ncb3300

***C. elegans* strains and culture conditions.** *C. elegans* strains were cultured under standard conditions[55]. Only hermaphrodites were used. A list of strains is available as Supplementary Table 8.

RNAi analysis. RNAi clones were derived from the Vidal RNAi library[56], the Ahringer RNAi library[57], or generated in-house by cloning of a PCR product amplified from cDNA into a modified L4440 RNAi feeding vector containing AscI and NotI restriction sites (generated by inserting linker L4440_link into BglII and KpnI digested L4440). A list of RNAi clones is available in Supplementary Table 9. All RNAi clones corresponding to bait proteins were sequence verified using primers L4440_F and L4440_R. All RNAi experiments were performed at 20 °C, except for experiments with strain JH2647, which was maintained at 25 °C to increase mCherry::PAR-6 expression levels. Bacteria were cultured overnight at 37 °C in lysogeny broth (LB) supplemented with 100 μg ml^{-1} ampicillin (Amp) and 2.5 μg ml^{-1} tetracycline (Tet). One hour before collection, 1 mM isopropyl β-D-1-thiogalactopyranoside (IPTG) was added. Bacterial cultures were concentrated 5× by centrifugation. Nematode growth media (NGM) agar plates supplemented with 100 μg ml^{-1} Amp, 2.5 μg ml^{-1} Tet, and 1 mM IPTG were seeded with 250 μl of bacterial suspension, and kept at room temperature for 48 h before use.

L4 hermaphrodites were placed on RNAi plates, and phenotypes were analysed in L4 stage F1 progeny, with exception of the mCherry::PAR-6 marker line, which was screened at the embryonic stage, and the yolk-trafficking strain, which was screened at the young adult stage. Larval phenotypes for genes that caused an embryonic lethal phenotype were examined by starting the RNAi at L1 and examining the L4 or adult stages of the same generation. Yolk-protein-trafficking phenotypes for genes that caused severe germline defects were examined by starting the feeding procedure in the L4 stage and analysing adults of the same generation.

Scoring of all RNAi experiments was done blind, with the observer not knowing the gene being analysed. In all experiments a positive RNAi control was included: *par-6* and *par-3* for early embryonic polarity, *pop-1* for the seam cell epithelium, *par-5* for the intestinal epithelium, *erm-1* for the excretory canal, *ajm-1* for epithelial junctions, *chc-1* and *cdc-42* for yolk-protein trafficking, *unc-33* for axon–dendrite specification, and *pop-1* for PVD neuron morphology. In the prey interactor RNAi screens, RNAi targeting the bait was always included, to compare the bait and prey RNAi phenotype and to confirm the originally observed bait defects. In all experiments, empty L4440 vector was used as a negative control. For each RNAi clone, 30–50 animals of the developmental stage indicated above were screened at ×10 magnification, and defects were characterized in detail and scored at ×63 or ×100 magnification. Severity of defects was scored on a scale of 0–5 on the basis of the number of worms affected and the strength of the effect in comparison with the positive control.

Fluorescence microscopy and live imaging. For live imaging of L3 and young adult hermaphrodites, animals were sedated with 10 mM levamisole M9 and mounted on 5% agarose. To perform imaging on embryos, gravid adults were dissected in M9 and mounted on 5% agarose. Two main microscopes were used. The first was a wide-field Zeiss Axioplan2 upright microscope equipped with ×25–0.8 NA, ×63–1.4 NA, and ×100–1.4 NA objectives, and an Axiocam MRm CCD (charge-coupled device) monochrome camera. Zeiss filter sets used were set 34 for DAPI, set 13 for GFP, and set 31 for mCherry. The microscope and camera were controlled by Zeiss Axiovision 4.x software. The second microscope was an Andor spinning-disc platform controlled by MetaMorph software and consisting of a Nikon Ti-U inverted microscope with ×60–1.4 NA and ×100–1.4 NA oil objectives, a Yokogawa CSU-X1 spinning-disc unit, 488 nm and 561 nm lasers, Semrock 512–23 + 630–91, 525–30, and 617–73 emission filters, and an Andor iXON DU-885 camera. Images in Fig. 6 were captured using a Zeiss AxioImager, ×40–1.3 NA objective, and a Hamamatsu Orca-R2 camera. Images of fixed embryos were deconvolved using AxioVision software, and are shown as maximum intensity projections of 3–5 adjacent planes spaced 0.3 μm apart. Images were cropped, rotated, and levels were adjusted in ImageJ and Adobe Photoshop.

Phenotype clustering. RNAi phenotypes were manually scored on a scale of 0 (no abnormal phenotype) to 5 (severely abnormal phenotype). For each gene, a binary phenotypic profile was generated consisting of the highest severity score for each of the 6 tissues analysed, adding defects in localization of junctional DLG-1 as a seventh data point. Clustering was done using the Cluster 3.0 software (http://bonsai.hgc.jp/~mdehoon/software/cluster/software.htm), using Spearman rank correlation distance measure and the complete linkage hierarchical clustering method. Clustering results were visualized using Java Treeview (http://sourceforge.net/projects/jtreeview).

Correlation of phenotype overlap with interaction, GO and PubMed. We examined all possible pairs of proteins in the network where the effect of inactivation had been determined in at least 1 overlapping tissue. Excluding homodimers, we evaluated 332 interacting protein pairs, and 8,373 pairs for which we did not detect a physical interaction. A phenotype was considered to overlap if both genes showed a detectable defect in at least 1 overlapping tissue. To examine the correlation between phenotype overlap and physical interaction, we compared the fraction of pairs with a phenotypic overlap between the 332 interacting protein pairs and the 8,373 non-interacting pairs. A Fisher's exact test was used to calculate the probability of the observed enrichment. To examine the correlation with GO terms, we used the HRSS scores described above. We defined a high similarity score as an HRSS score >0.9, and a low similarity score as an HRSS score ≤ 0.1. A Fisher's exact test was used to calculate the probability of observed enrichments. To examine the correlation between phenotype overlap and prior characterization of an interaction in a publication, we examined the 19 PubMed papers listed in Supplementary Table 3, excluding homodimers.

Yeast two-hybrid analysis of PAC-1–PAR-6. Fragments indicated in Fig. 6c were cloned into the Y2H vectors pMB28 and pMB29, and transformed into yeast strains Y8930 and Y8800 respectively. Yeast pairs were then mated on YEPD plates, transferred to –Leu –Trp plates for selection of mated yeast, and transferred to –Leu –Trp –His plates to assay the interaction.

Generation of *pac-1* transgenic lines. A PAC-1 Gateway clone lacking amino acids 1221–1328 was produced by PCR-modification of a full-length *pac-1* cDNA entry clone[45], and then recombined into destination vector pID3.01B (ref. 58) to create fusions with *GFP*. P*pie-1::GFP::pac-1(ΔPBD)* was integrated into *unc-119(ed3)* worms using biolistic transformation[59].

Antibody staining of embryos. Embryos were freeze-fractured, fixed in methanol, and stained as described previously[45]. The following primary antibodies were used: rabbit α-PAR-6 1:20,000 (ref. 60), rat α-GFP 1:1,000 (Nacalai Tesque GF090R).

Quantifying polarity index. Polarity index was quantified in 8–12-cell-stage embryos. Polarity index measurements were obtained by determining the ratio of the average intensity of a 40-pixel line along the contact-free surface of an AB lineage cell versus half of the average intensity of a 40-pixel line at that cell's contact with a neighbouring AB cell. ImageJ was used for intensity measurements.

Quantifying expression levels of GFP::PAC-1 and GFP::PAC-1(ΔPBD). Total fluorescence values were obtained in a central focal plane of 4-cell embryos expressing each GFP fusion. Using equivalent camera exposures, average background fluorescence (wild-type 4-cell embryos) was subtracted from the average fluorescence value for each strain. Signals were integrated over the area of the embryo.

Co-localization in HeLa cells. Avi–mCherry–C1 and pEGFP–C1 constructs generated for co-affinity purification were transfected in HeLa Kyoto cells grown in DMEM-Ham's F10 (50/50%) medium containing 10% FCS and 1% penicillin + streptomycin at 37 °C and 5% CO_2. Two days before transfection, nearly confluent cells were plated at 1:10 in 12-well plates on 24 mm glass coverslips. Cells were transfected with 1 μg of Avi–mCherry–bait and 1 μg of pEGFP–C1–prey with polyethyleminine (PEI). After overnight incubation, cells were washed with PBS, fixed for 10 min with 4% paraformaldehyde in PBS at room temperature, washed with PBS and subsequently mounted on slides in Vectashield mounting medium with DAPI. The original source of the HeLa Kyoto cells is not known and cells were not authenticated.

Statistical analysis. Statistical tests used and sample sizes are indicated in the figure legends and in the Methods section. No statistical method was used to predetermine sample sizes. No samples or animals were excluded from analysis. The experiments were not randomized, and the investigators were not blinded to allocation during experiments and outcome assessment, with the exception of the RNAi screens, where investigators were unaware of the gene being targeted during scoring. Statistics regarding the enrichment of protein pairs in the interaction network sharing certain characteristics were based on comparisons with random control networks, with P values corresponding to the number of times a value as extreme as the actual network was found in 100,000 control networks. This statistical approach was chosen to reduce bias by factors such as the connectivity degree of the proteins in the network. A two-sided Fisher's exact test was used to evaluate the significance of a higher fraction of interacting pairs retesting positively by MAPPIT than control random pairs. The test is appropriate to examine whether there is a significant association between the retest result by MAPPIT and prior identification as an interacting protein pair by an interaction assay. A Mann–Whitney U-test was used to determine the significance in differences in polarity index and fluorescence intensity levels between embryos of different genotypes, as a normal distribution of the data could not be confidently determined from the sample values (Anderson–Darling normality test).

Reproducibility of experiments. Co-affinity purifications in Fig. 2d and Supplementary Fig. 2 were performed once. Localization of fluorescent marker proteins in wild-type background (Fig. 4a–i) was observed in >30 animals. All RNAi phenotypes (Fig. 5 and Supplementary Table 6) were observed in >5 animals. Y2H analyses in Fig. 6c were performed three times. Co-affinity purifications in Fig. 6d were performed twice. PAC-1 and PAR-6 localization patterns in HeLa cells (Fig. 6e–g) were observed in >30 cells. Embryo staining patterns in Fig. 6h–k were observed in 58/58 (h), 62/62 (i), 50/50 (j) and 61/61 (k) embryos.

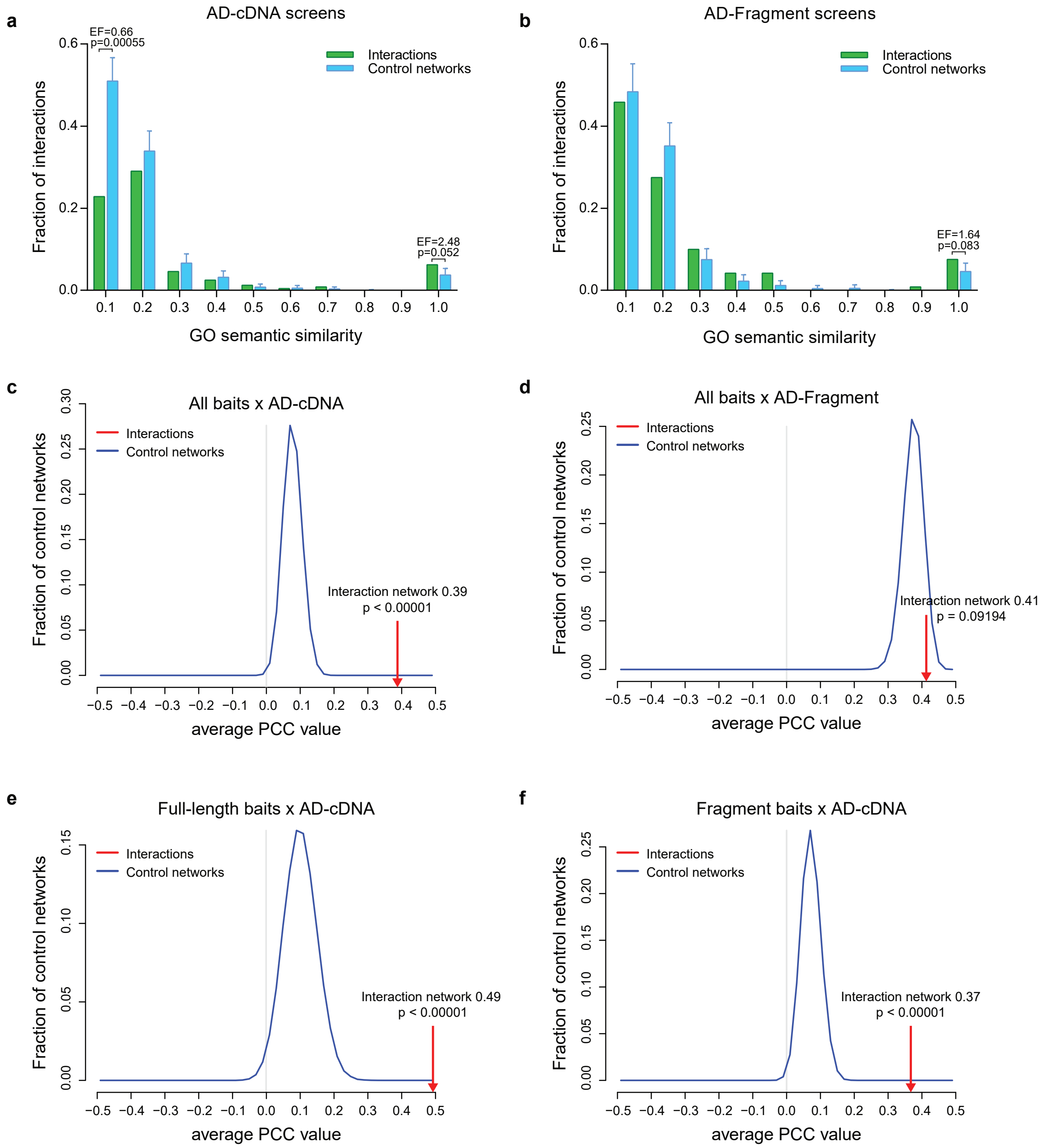
Reanalysis of published data sets. In Fig. 2e, we reanalysed the MAPPIT results from three *C. elegans* protein interaction data sets previously analysed by MAPPIT (ref. 18).

Code availability. All code that was generated in-house is available on request. Code may not be further distributed or used for purposes other than to reanalyse the data presented here without prior permission from the authors.

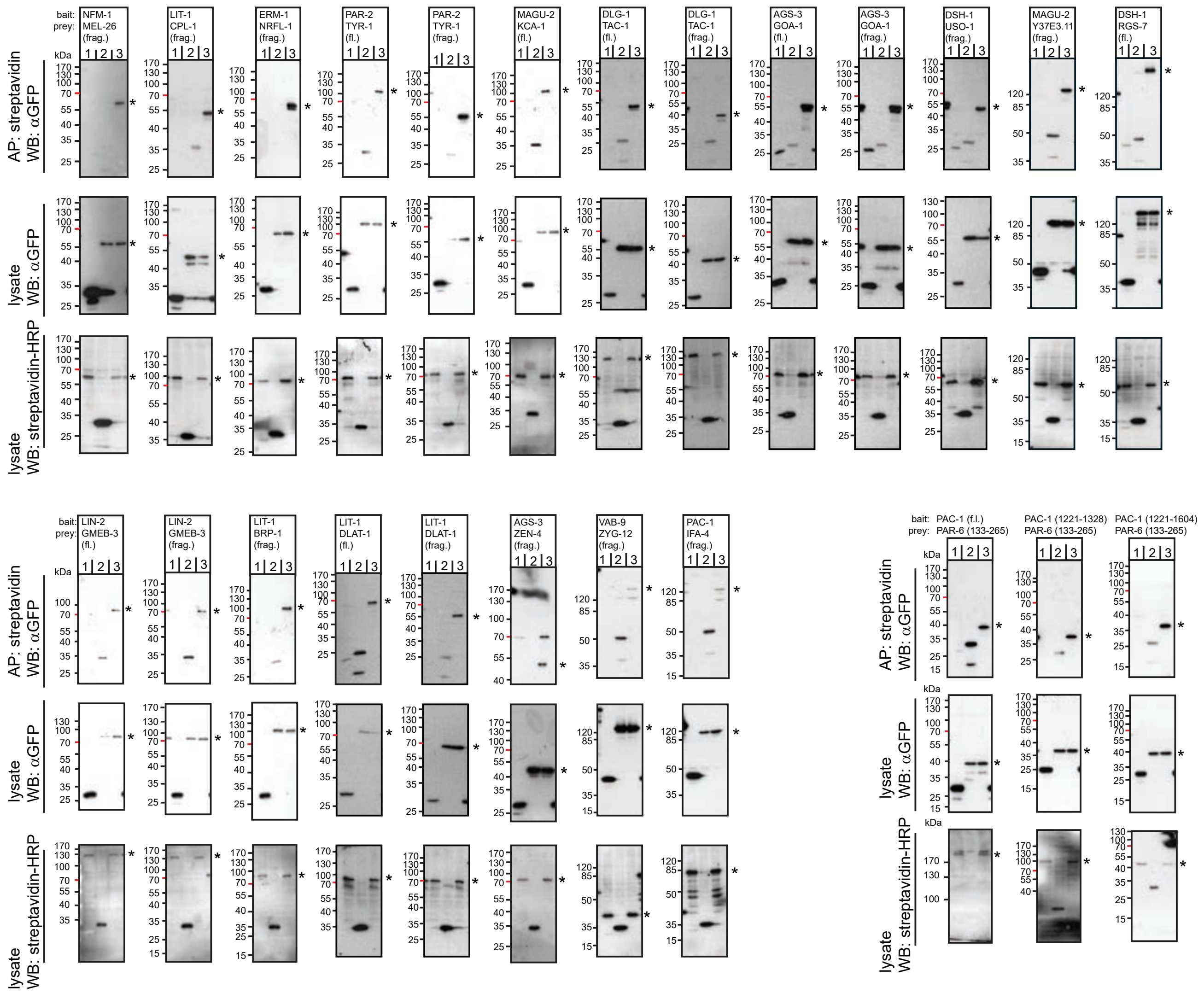
Data availability. The protein interactions from this publication have been submitted to the IMEx (http://www.imexconsortium.org) consortium through the IntAct database[61] and assigned the identifier IM-24524. A searchable web interface of all interactions and fragments identified is available at http://www.projects.science.uu.nl/interactome.

47. Chevray, P. M. & Nathans, D. Protein interaction cloning in yeast: identification of mammalian proteins that react with the leucine zipper of Jun. *Proc. Natl Acad. Sci. USA* **89,** 5789–5793 (1992).
48. Vidal, M., Brachmann, R. K., Fattaey, A., Harlow, E. & Boeke, J. D. Reverse two-hybrid and one-hybrid systems to detect dissociation of protein–protein and DNA-protein interactions. *Proc. Natl Acad. Sci. USA* **93,** 10315–10320 (1996).
49. Yu, H. *et al.* High-quality binary protein interaction map of the yeast interactome network. *Science* **322,** 104–110 (2008).
50. Schiestl, R. H. & Gietz, R. D. High efficiency transformation of intact yeast cells using single stranded nucleic acids as a carrier. *Curr. Genet.* **16,** 339–346 (1989).
51. Ewing, B., Hillier, L., Wendl, M. C. & Green, P. Base-calling of automated sequencer traces using phred. I. Accuracy assessment. *Genome Res.* **8,** 175–185 (1998).
52. Hibbs, M. A. *et al.* Exploring the functional landscape of gene expression: directed search of large microarray compendia. *Bioinformatics* **23,** 2692–2699 (2007).
53. R Core Team *R: A Language and Environment for Statistical Computing* (R Foundation for Statistical Computing, 2014); http://www.R-project.org
54. Lin, Y.-C. *et al.* Genome dynamics of the human embryonic kidney 293 lineage in response to cell biology manipulations. *Nat. Commun.* **5,** 4767 (2014).
55. Brenner, S. The genetics of *Caenorhabditis elegans*. *Genetics* **77,** 71–94 (1974).
56. Lamesch, P. *et al. C. elegans* ORFeome version 3.1: increasing the coverage of ORFeome resources with improved gene predictions. *Genome Res.* **14,** 2064–2069 (2004).
57. Fraser, A. G. *et al.* Functional genomic analysis of *C. elegans* chromosome I by systematic RNA interference. *Nature* **408,** 325–330 (2000).
58. D'Agostino, I., Merritt, C., Chen, P.-L., Seydoux, G. & Subramaniam, K. Translational repression restricts expression of the *C. elegans* Nanos homolog NOS-2 to the embryonic germline. *Dev. Biol.* **292,** 244–252 (2006).
59. Praitis, V. Creation of transgenic lines using microparticle bombardment methods. *Methods Mol. Biol.* **351,** 93–107 (2006).
60. Schonegg, S. & Hyman, A. A. CDC-42 and RHO-1 coordinate acto-myosin contractility and PAR protein localization during polarity establishment in *C. elegans* embryos. *Development* **133,** 3507–3516 (2006).
61. Orchard, S. *et al.* The MIntAct project–IntAct as a common curation platform for 11 molecular interaction databases. *Nucleic Acids Res.* **42,** D358–D363 (2014).

DOI: 10.1038/ncb3300



Supplementary Figure 1 Enrichment of shared Gene Ontology (GO) terms and similar expression profiles in interacting protein pairs compared to control networks generated by replacing the prey proteins with random proteins from the search space. (**a**) GO similarity scores for interacting protein pairs identified using the AD-cDNA library. Semantic similarity scores for the BP component of GO were calculated using the HRSS software package, which scores protein pairs on a scale of 0 (dissimilar GO terms) to 1 (similar GO terms)[29]. (**b**) Same as (**a**), but for interactions identified using the AD-Fragment scores. Due to the already high semantic similarity scores for pairs in the search space (749 genes involved in early embryonic development), enrichment scores are less high than for AD-cDNA derived interactions. Control network bars represent the mean of 100,000 control networks ± s.d. Statistical significance is the fraction of control networks that displayed the same or higher fraction of pairs with a particular GO similarity range as the actual interaction network. Protein pairs in the interaction network are depleted for pairs with a low semantic similarity score, and enriched for pairs with a high similarity score. (**c**–**f**) Average Pearson correlation coefficient (PCC) score of mRNAs corresponding to protein pairs in the indicated interaction datasets (red arrows), compared with the distribution of average PCC scores of 100,000 control networks generated by replacing the bait proteins with random proteins from the search space (blue line). PCC values were calculated using the compendium of expression microarray data collected in Wormbase release WS236. Statistical significance is the fraction of control networks that displayed an average PCC score identical or higher than the actual interaction network. As observed previously[9], early embryogenesis genes already have such similar expression profiles that no further enrichment can be observed for interactions derived from the AD-Fragment library.

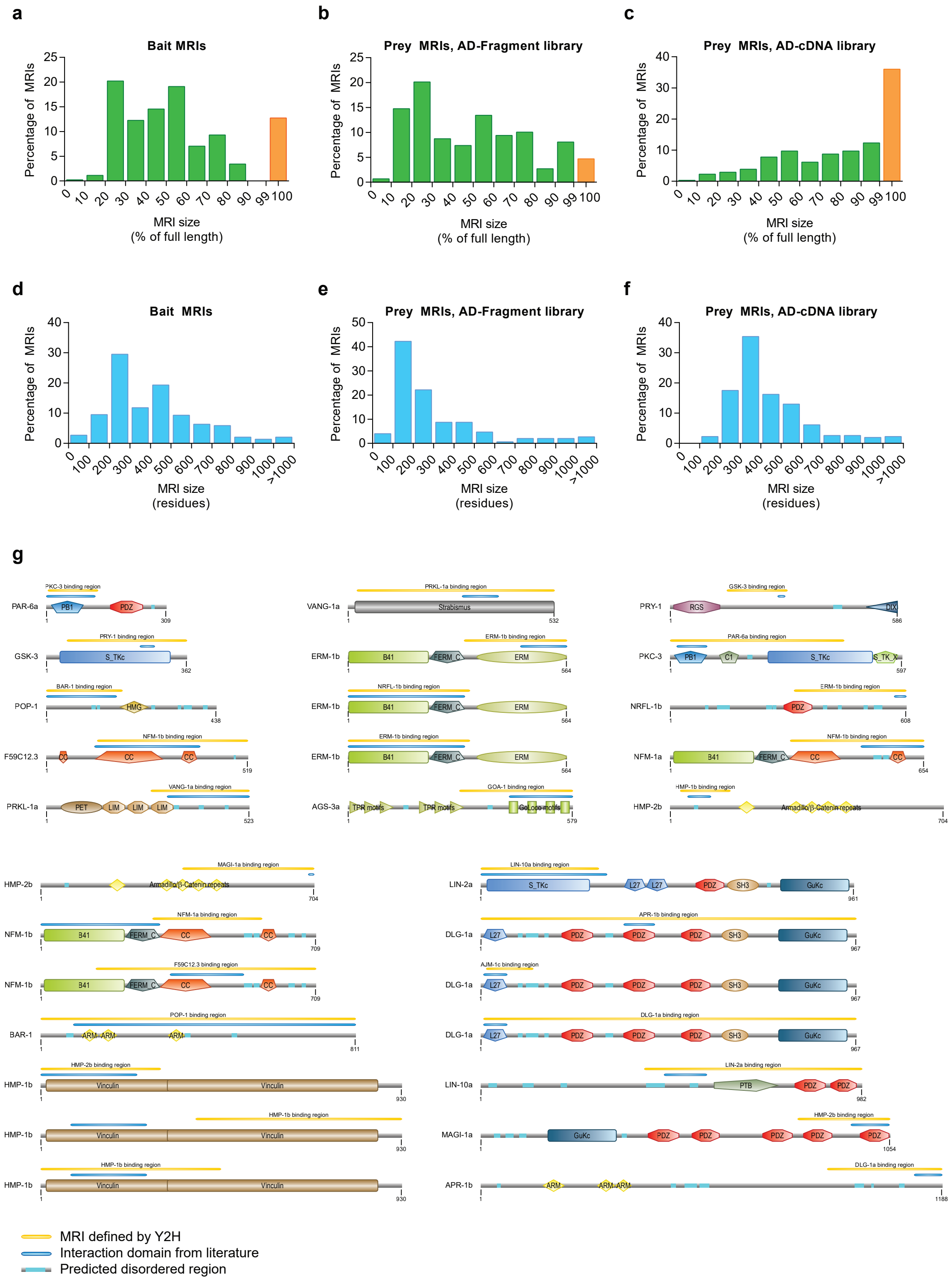


Lanes:

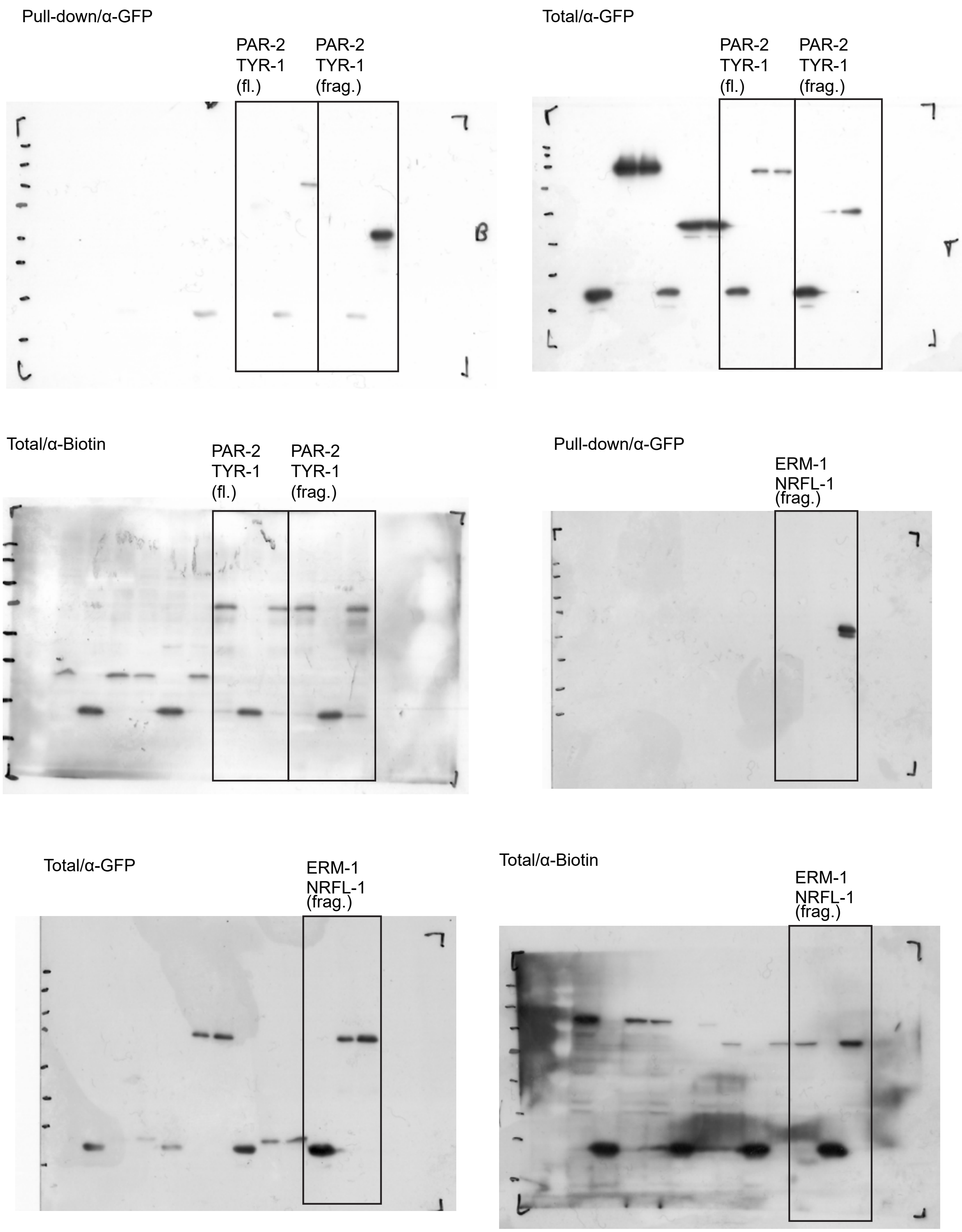
1: EGFP + Avi-mCherry-bait
2: EGFP-prey + Avi-mCherry
3: EGFP-prey + Avi-mCherry-bait

Supplementary Figure 2 Western blots for all protein pairs testing positive by co-affinity purification. For every pair, three blots are shown. Top: detection of GFP-tagged proteins that co-purify with biotinylated mCherry-tagged proteins using an anti-GFP antibody. Middle: detection of GFP-tagged proteins in total lysates using an anti-GFP antibody. Bottom: detection of biotinylated Avi-mCherry tagged proteins in total lysates using streptavidin coupled to horse radish peroxidase. Protein pairs tested are indicated above the blots. The first protein listed is the Avi-mCherry tagged bait. Also indicated is whether the prey protein tested was full-length (f.l.) or corresponds to a shorter fragment (frag.). In all blots, lanes 1 and 2 are negative controls, and lane 3 is the actual affinity purification (1: Avi-mCherry-bait vs. EGFP alone, 2: Avi-mCherry alone vs. EGFP-tagged prey, 3: Avi-mCherry-bait vs. EGFP-prey). The ~35 kDa band in lanes 2 of the upper blots is due to cross-reactivity of the anti-GFP antibody with the highly abundant Avi-mCherry polypeptide. Asterisks indicate bands of expected molecular mass. Purifications were performed once. Unprocessed scans are shown in Supplementary Figure 5.

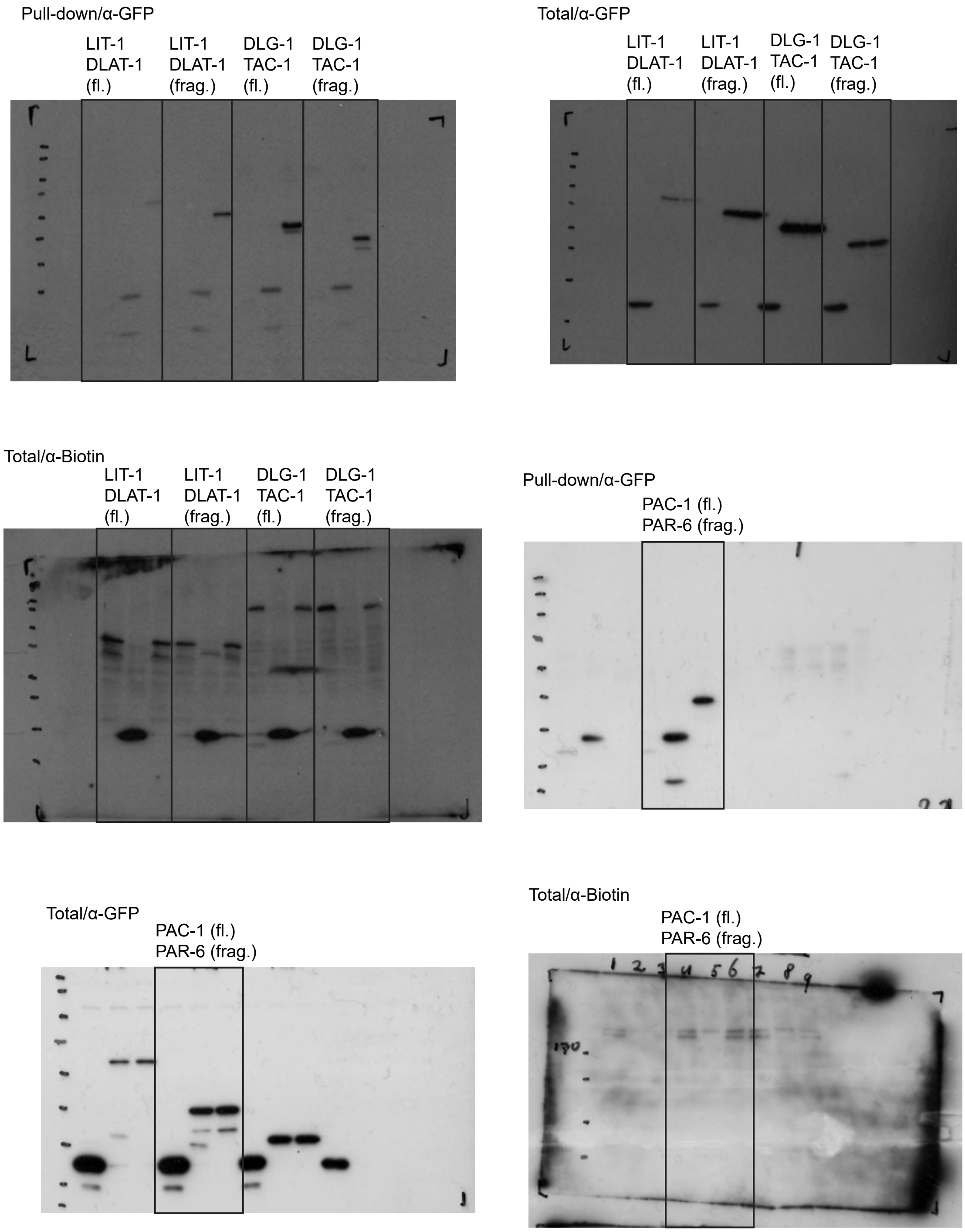
Supplementary Figure 3 Graphical representation of every protein in the interaction network and the MRIs identified for that protein. Grey boxes are full-length proteins. Predicted domains are shown as boxes of various colors and shapes. Identified MRIs are shown as yellow lines above the protein graphic. Protein names above each MRI indicate which proteins bind to that particular region.



Supplementary Figure 4 Analysis of MRIs (**a**–**c**) Distributions of MRI lengths expressed as the percentage of the corresponding full-length protein. (**d**–**f**) Distributions of absolute MRI lengths in amino acid residues. (**a**,**d**) MRIs identified for the bait proteins. (**b**,**e**) MRIs identified for the prey proteins from AD-Fragment library derived clones. (**c**,**f**) MRIs identified for the prey proteins from AD-cDNA library derived clones. (**g**) Graphical representations of all MRIs where interaction sites had previously been identified in the literature. Grey boxes are full-length proteins. Predicted domains are shown as boxes of various colors and shapes. Identified MRIs are shown as yellow lines above the protein graphic. Interaction sites from the literature are shown as blue lines above the protein graphic. In cases where the literature describes an interaction domain for a non-*C. elegans* protein, the corresponding site in the *C. elegans* protein is shown.



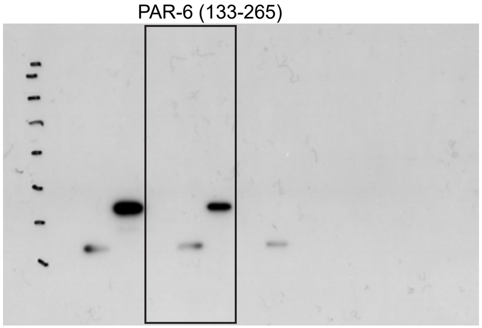
Supplementary Figure 5 Unprocessed scans of the blots shown in Figure 2, Figure 6, and Supplementary Figure 2.



Supplementary Figure 5 continued

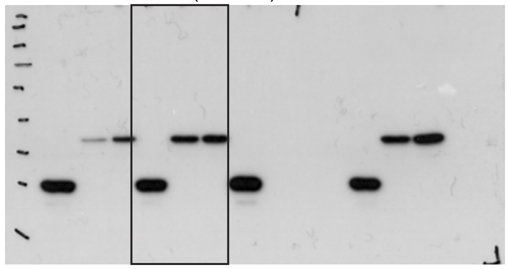
Pull-down/α-GFP

PAC-1 (1221-1604)
PAR-6 (133-265)



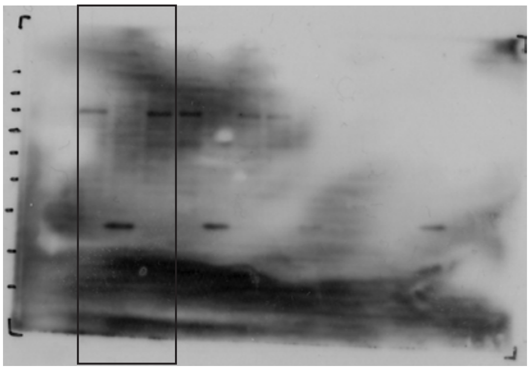
Total/α-GFP

PAC-1 (1221-1604)
PAR-6 (133-265)



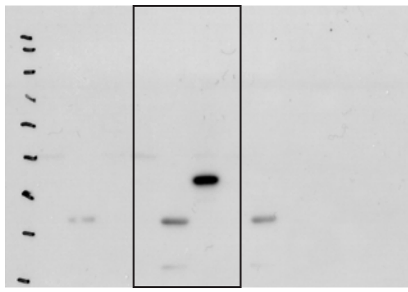
Total/α-Biotin

PAC-1 (1221-1604)
PAR-6 (133-265)



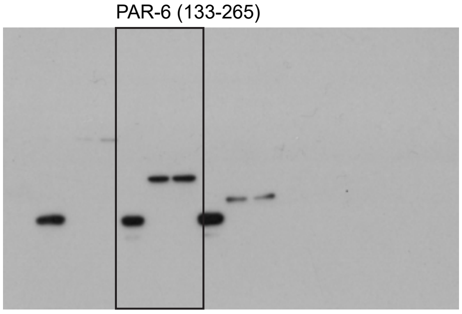
Pull-down/α-GFP

PAC-1 (1221-1328)
PAR-6 (133-265)



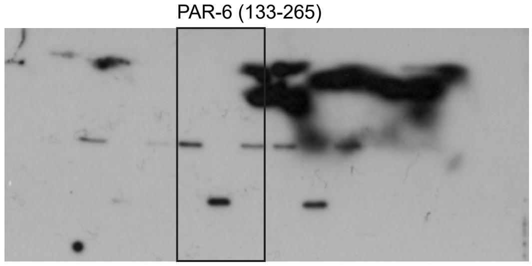
Total/α-GFP

PAC-1 (1221-1328)
PAR-6 (133-265)

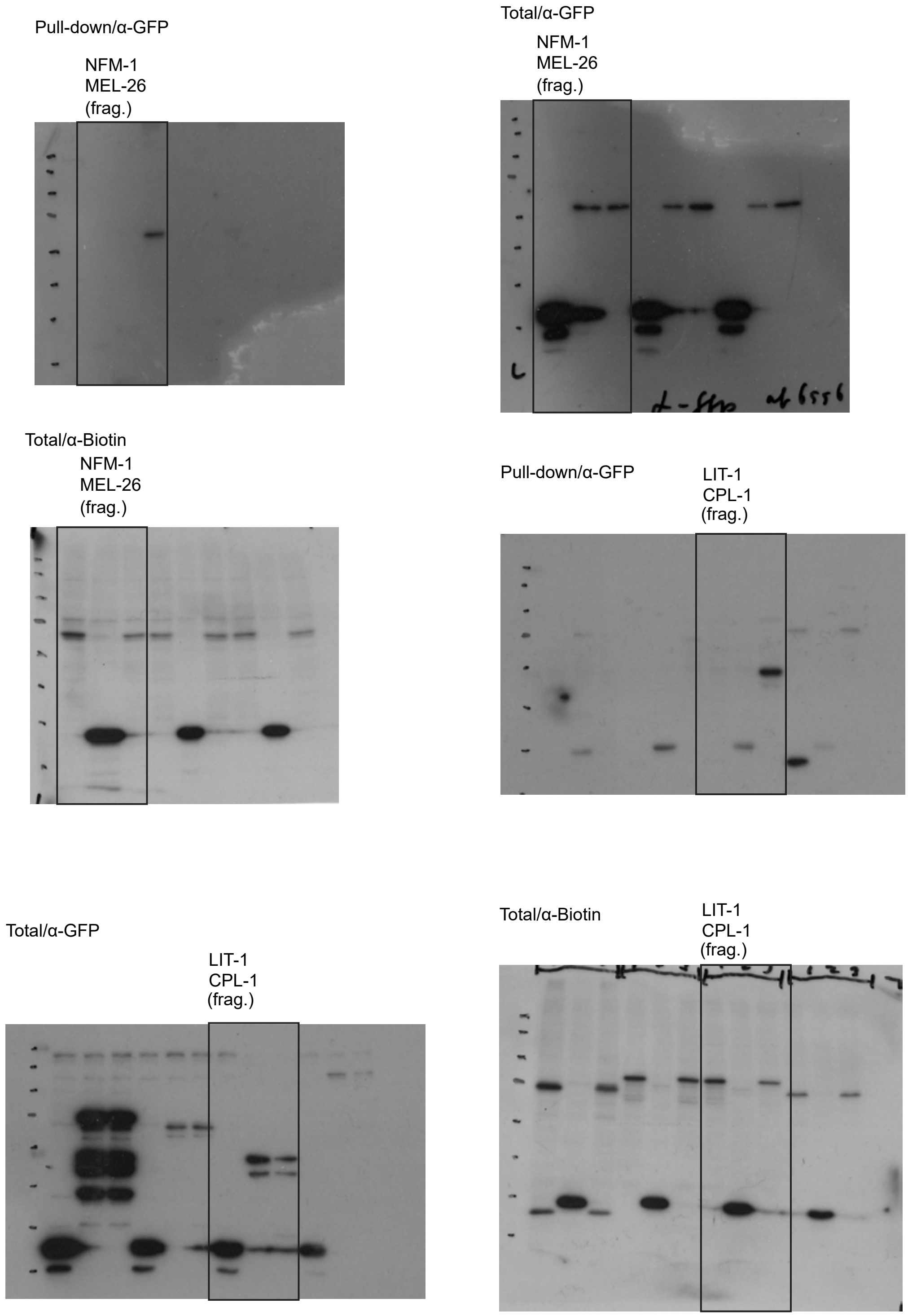


Total/α-Biotin

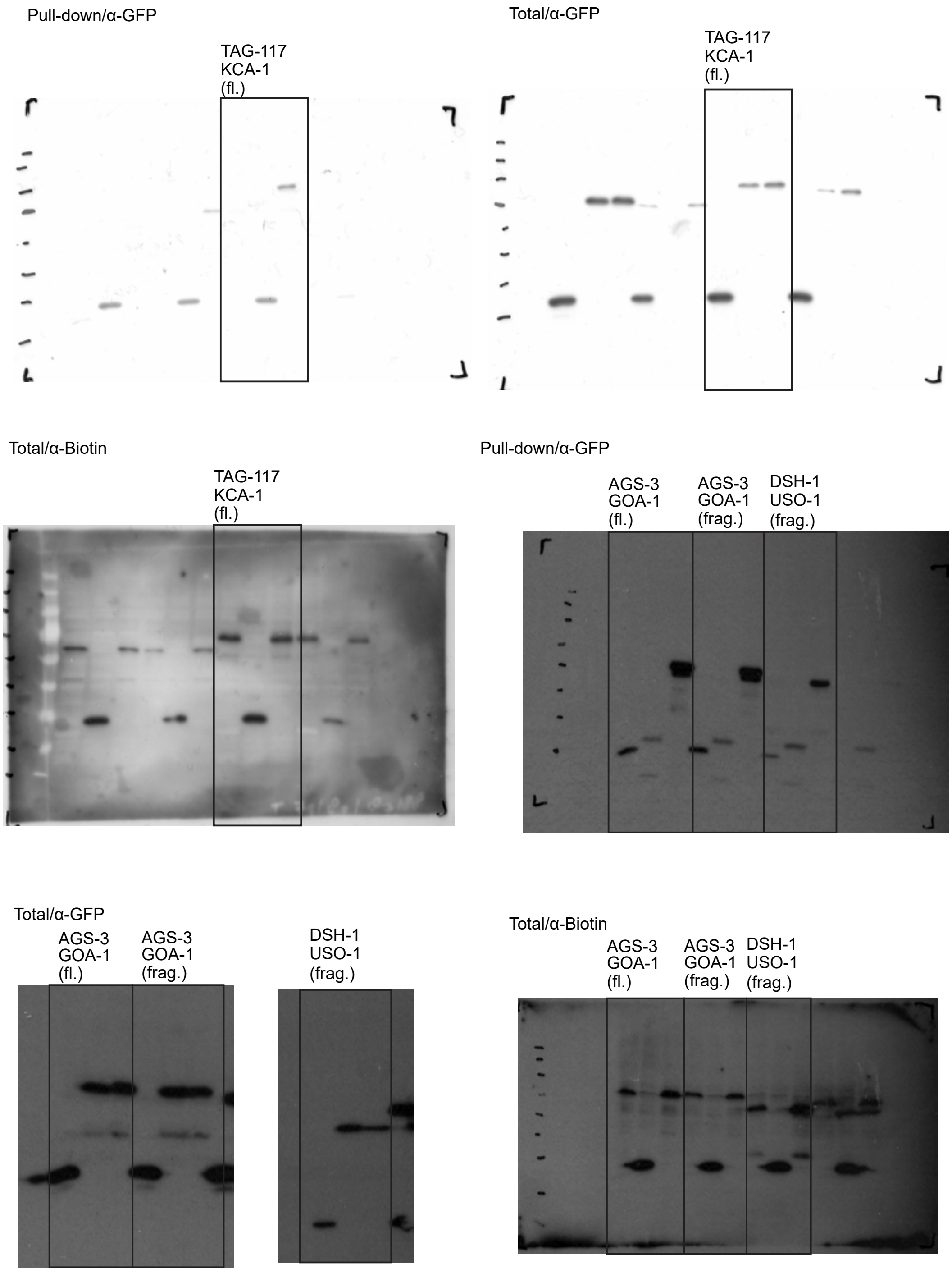
PAC-1 (1221-1328)
PAR-6 (133-265)



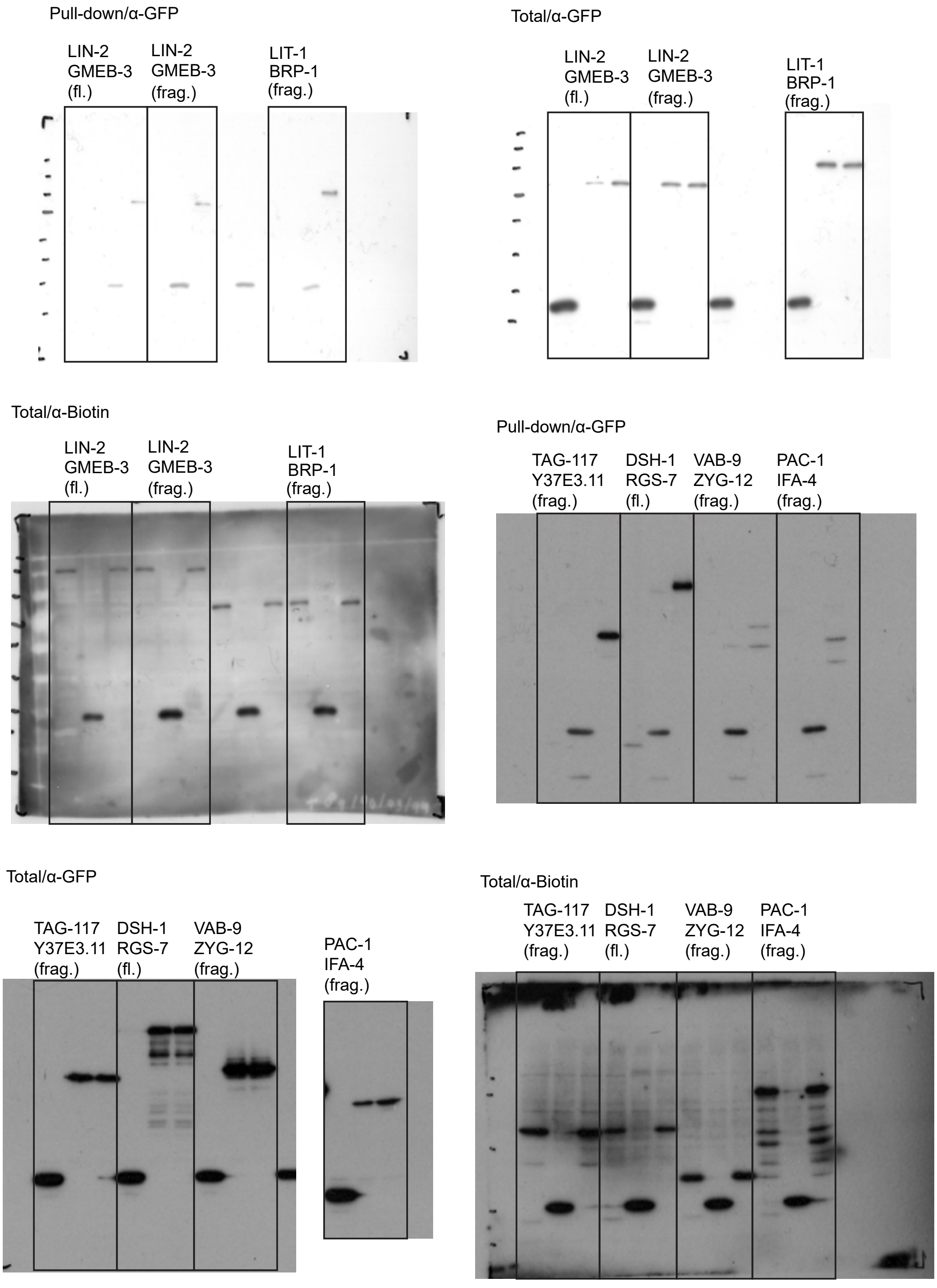
Supplementary Figure 5 continued



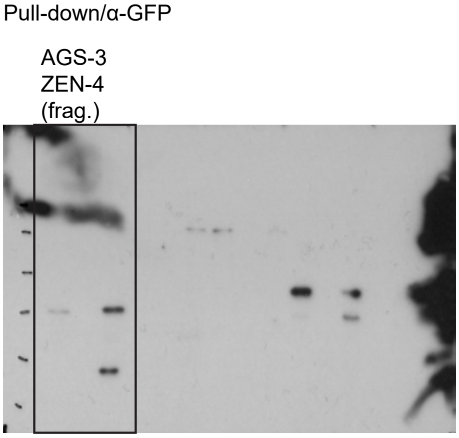
Supplementary Figure 5 continued

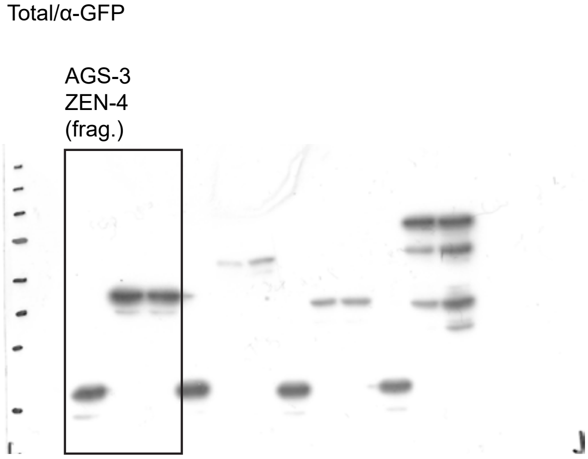


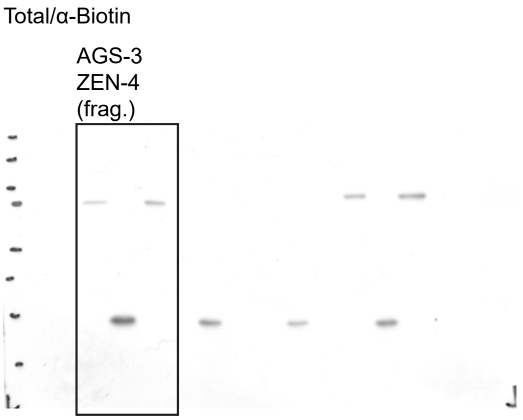
Supplementary Figure 5 continued



Supplementary Figure 5 continued

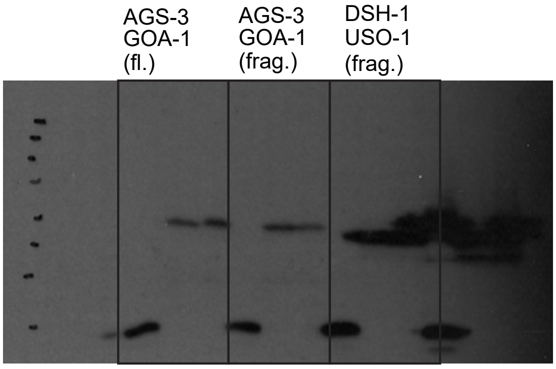


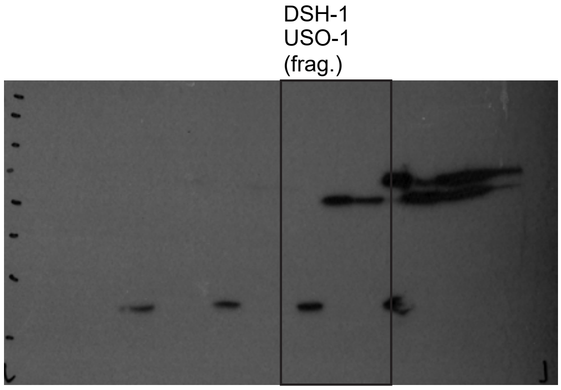




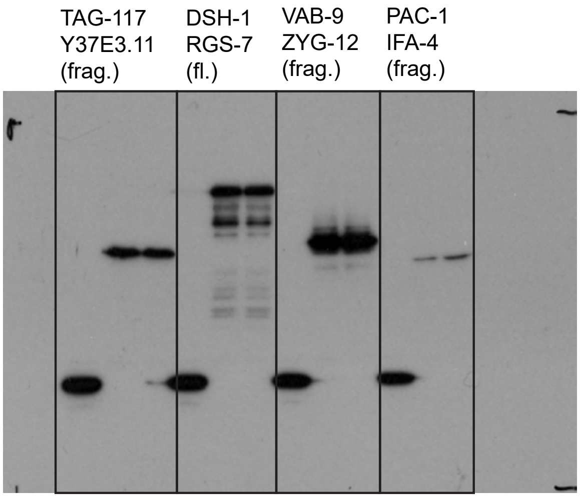
Supplementary Figure 5 continued

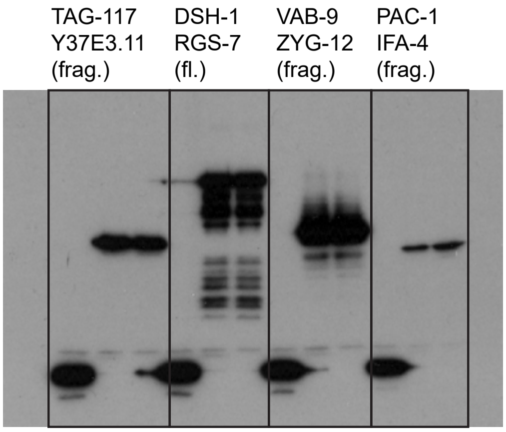
Total/α-GFP: The two blots displayed here are the uncropped versions of the cropped blots on page 5





Total/α-GFP: The two blots displayed here are the uncropped versions of the cropped blots on page 6





Supplementary Figure 5 continued

Supplementary Table Legends

Supplementary Table 1 Bait protein and bait clone information. (Sheet 1) Bait proteins used in this study, and homologs in other organisms. (Sheet 2) Summary of the cloning and screening success rates for each bait protein. (Sheet 3) Details of the bait clones generated, including start and end coordinates. (Sheet 4) Domain predictions for the bait proteins.

Supplementary Table 2 The *C. elegans* polarity interaction network. (Sheet 1) The full interaction network, including information on which protein fragments were identified, and how often each interaction was identified. (Sheet 2) Interactions identified using full-length bait constructs. (Sheet 3) Interactions identified using fragment bait constructs.

Supplementary Table 3 Comparison of the polarity interaction network with literature descriptions of protein interactions. (Sheet 1) Overlap between the *Ce*PIN and interactions between *C. elegans* proteins present in the IntAct database. (Sheet 2) Overlap between the *Ce*PIN and interactions between homologous proteins present in the IntAct database. (Sheet 3) Previously published interactions from the *Ce*PIN identified through manual PubMed searches. (Sheet 4) Summary of all overlaps in sheets 1-3.

Supplementary Table 4 Validation of the polarity interaction network by affinity purification and MAPPIT. (Sheet 1) Results of retest by affinity purification from HEK-293 cells. (Sheet 2) Results of retest by MAPPIT.

Supplementary Table 5 Minimal Regions of Interactions (MRIs) identified. (Sheet 1) MRIs for bait and prey protein for each interaction in the *Ce*PIN. (Sheet 2) Comparison of MRIs with interaction domain information in the literature. (Sheet 3) List of the fragments from which the MRIs for the bait proteins were delineated. (Sheet 4) List of the fragments from which the MRIs for the prey proteins were delineated.

Supplementary Table 6 Phenotypic analysis by RNAi. (Sheet 1) List of the marker strains analyzed and phenotypes scored. (Sheet 2) List of genes inactivated and phenotypes observed. (Sheet 3) Summary of sheet 2, showing the strongest phenotype observed in each of the tissues analyzed. (Sheet 4) List of interacting protein pairs where RNAi for both corresponding genes results in a phenotype in the same tissue(s). (Sheet 5) Detailed information on the phenotypes observed for the interactions listed on sheet 4. (Sheet 6) Phenotype overlap between *Ce*PIN interactions that were previously published.

Supplementary Table 7 List of primers used in this study. (Sheet 1) Primers used to generate the bait clones. (Sheet 2) Primers used to generate clones for affinity purification and expression in HeLa cells. (Sheet 3) Primers used to generate clones for MAPPIT. (Sheet 4) Primers used to generate RNAi clones. (Sheet 5) All other primers.

Supplementary Table 8 List of *C. elegans* strains used in this study.

Supplementary Table 9 Source of the RNAi clones used in this study.