Transcription factor modularity in a gene-centered *C. elegans* core neuronal protein–DNA interaction network

Vanessa Vermeirssen, M. Inmaculada Barrasa, César A. Hidalgo, et al.

Supplemental Material	http://genome.cshlp.org/content/suppl/2007/05/21/gr.6148107.DC1.html
References	This article cites 73 articles, 29 of which can be accessed free at: http://genome.cshlp.org/content/17/7/1061.full.html#ref-list-1
Creative Commons License	This article is distributed exclusively by Cold Spring Harbor Laboratory Press for the first six months after the full-issue publication date (see http://genome.cshlp.org/site/misc/terms.xhtml). After six months, it is available under a Creative Commons License (Attribution-NonCommercial 3.0 Unported License), as described at http://creativecommons.org/licenses/by-nc/3.0/.
Email Alerting Service	Receive free email alerts when new articles cite this article - sign up in the box at the top right corner of the article or **click here.**

Letter

Transcription factor modularity in a gene-centered *C. elegans* core neuronal protein–DNA interaction network

Vanessa Vermeirssen,[1] M. Inmaculada Barrasa,[1] César A. Hidalgo,[2]
Jenny Aurielle B. Babon,[1] Reynaldo Sequerra,[3] Lynn Doucette-Stamm,[3]
Albert-László Barabási,[2] and Albertha J.M. Walhout[1,4]

[1]*Program in Gene Function and Expression and Program in Molecular Medicine, University of Massachusetts Medical School, Worcester, Massachusetts 01605, USA;* [2]*Center for Complex Network Research, Department of Physics, University of Notre Dame, Notre Dame, Indiana 46556, USA;* [3]*Agencourt Bioscience Corporation, Beverly, Massachusetts 01915, USA*

Transcription regulatory networks play a pivotal role in the development, function, and pathology of metazoan organisms. Such networks are comprised of protein–DNA interactions between transcription factors (TFs) and their target genes. An important question pertains to how the architecture of such networks relates to network functionality. Here, we show that a *Caenorhabditis elegans* core neuronal protein–DNA interaction network is organized into two TF modules. These modules contain TFs that bind to a relatively small number of target genes and are more systems specific than the TF hubs that connect the modules. Each module relates to different functional aspects of the network. One module contains TFs involved in reproduction and target genes that are expressed in neurons as well as in other tissues. The second module is enriched for paired homeodomain TFs and connects to target genes that are often exclusively neuronal. We find that paired homeodomain TFs are specifically expressed in *C. elegans* and mouse neurons, indicating that the neuronal function of paired homeodomains is evolutionarily conserved. Taken together, we show that a core neuronal *C. elegans* protein–DNA interaction network possesses TF modules that relate to different functional aspects of the complete network.

[Supplemental material is available online at www.genome.org.]

Differential gene expression is an important driving force in the development, function, and pathology of multicellular organisms. Differential gene expression is first regulated at the level of transcription initiation by regulatory transcription factors (TFs) that directly bind to their genomic DNA targets, resulting in an activation or repression of target gene expression. In metazoans, 5%–10% of the genes encode predicted TFs (Levine and Tjian 2003; Reece-Hoyes et al. 2005), each of which likely regulates the expression of multiple target genes. TFs function in the context of intricate transcription regulatory networks that describe gene expression as a function of inputs specified by physical and functional interactions between TFs and DNA (for review, see Blais and Dynlacht 2005; Davidson and Levine 2005; Walhout 2006). Although transcription regulatory networks in unicellular systems have been studied extensively, the architecture and functionality of the networks that control multicellular development and function remain poorly understood.

A first step in deciphering transcription regulatory networks is the large-scale mapping of protein–DNA interactions (PDIs) between TFs and their target genes (Walhout 2006). TF-centered approaches such as chromatin immunoprecipitation (ChIP), can be used to identify the DNA sequences bound by a TF in vivo (for review, see Blais and Dynlacht 2005; Elnitski et al. 2006). In complex metazoan systems, such methods are limited to TFs that are widely and highly expressed, and for which suitable antibodies

are available. Gene-centered methods such as the Gateway-compatible yeast one-hybrid system (Y1H) provide a high-throughput, condition-independent approach for the systematic identification of PDIs between gene promoters and TFs (Deplancke et al. 2004). The Y1H system can be used to efficiently identify a wide variety of metazoan TFs. For instance, we recently described a *Caenorhabditis elegans* digestive tract PDI network containing >100 predicted TFs, most of which were previously uncharacterized (Deplancke et al. 2006a). In addition to validating multiple Y1H interactions in vivo, we demonstrated that this PDI network is enriched for TFs that are themselves expressed in the digestive tract. This suggests that the gene-centered Y1H system enables the identification of specific PDI networks that involve genes and TFs that function in a tissue of interest.

In *C. elegans,* 15 types of neurons sense the chemical environment or temperature (Melkman and Sengupta 2005). These neurons are defined by the combinatorial expression of terminal differentiation genes, or "gene batteries." This expression is accomplished through the action of different combinations of TFs (Hobert 2005). Several TFs have been reported to determine neuronal cell fate and function (Lanjuin and Sengupta 2004; Hobert 2005). However, it is unclear how the expression of these TFs themselves is regulated. To gain insight into the transcription regulatory networks that govern neuronal TF expression, we mapped a PDI network with all TF-encoding genes that are known to be expressed and/or function in *C. elegans* chemo- and thermosensory neurons. Since only TF-encoding genes are used as target genes, the resulting network can be considered a "core

[4]**Corresponding author.**
E-mail marian.walhout@umassmed.edu; fax (508) 856-5460.

neuronal PDI network" (Davidson et al. 2003; Deplancke et al. 2006a).

We find that the core neuronal PDI network contains two TF modules that each associate with different functional aspects of the network: one module contains TFs involved in reproduction and connects to target genes that are expressed in neurons and other tissues, whereas the other module is enriched for paired homeodomain TFs that bind to target genes that are primarily expressed in neurons. We find that paired homeodomain TFs tend to be exclusively expressed in neurons in both *C. elegans* and in mouse, suggesting that their neuronal function is evolutionarily conserved.

Results

Mapping a core neuronal PDI network by gene-centered Y1H assays

We selected 50 promoters of TF-encoding genes as DNA baits for Y1H assays. These promoters correspond to TF-encoding genes that are known to be expressed and/or function in chemo- and thermosensory neurons and their interneurons (Supplemental Table S1). We successfully created Y1H bait strains for 47 of the 50 promoters (94%, Supplemental Fig. S1; Supplemental Table S2). To attain high PDI coverage, we performed four Y1H assays: all baits were screened versus an AD-wrmcDNA library (Walhout et al. 2000b) and an AD-TF mini-library (Deplancke et al. 2004), several baits (Supplemental Table S2) were mated versus an AD-TF yeast array (V. Vermeirssen, B. Deplancke, M.I. Barrasa, J.S. Reece-Hoyes, H.E. Arda, C.A. Grove, N.J. Martinez, R. Sequerra, L. Doucette-Stamm, M.R. Brent, et al., in prep.) and all baits were used in a final Y1H matrix experiment using available interactor TFs (Supplemental Fig. S1; Supplemental Table S2). Combined, these assays retrieved 376 PDIs (Supplemental Table S3).

High-throughput methods have the advantage to rapidly generate large datasets, but such datasets may contain false positive information. In order to minimize the inclusion of false positives, we developed a stringent standardized scoring system for high-throughput Y1H assays and applied it to our data set (Supplemental Materials; Supplemental Fig. S2; Supplemental Table S4). The scoring system takes several criteria, which contribute to the quality of a PDI into account, including the DNA bait, the interactor prey, and the interaction itself. After scoring, we extracted a high-confidence Y1H data set consisting of 282 PDIs between 38 promoters and 94 interactors. We visualized the high-confidence PDIs into a "core neuronal PDI network" graph. All PDIs, except one, were connected into a single network (Fig. 1A; Supplemental Fig. S4). All PDIs can be accessed in our EDGEdb database (Barrasa et al. 2007, http://edgedb.umassmed.edu).

From PDIs to transcription regulatory interactions by data integration

Y1H data do not provide insight into the transcriptional consequences of PDIs, i.e., activation or repression (Walhout 2006). By integrating physical interactions with previously reported regulatory information, we converted several PDIs into transcription regulatory interactions (Fig. 1B). We detected a total of seven autoregulatory loops, i.e., TFs binding their own promoter, involving CEH-10, DAC-1, DAF-19, NHR-41, ODR-7, UNC-30, and UNC-42. Three of these confirm physical interactions for previously suggested autoregulatory events (Fig. 1B) (Baran et al. 1999; Swoboda et al. 2000; Sarafi-Reinach et al. 2001).

We found two novel physical interactions between components that were known to function together in the diversification of bilateral asymmetric neurons: DIE-1 interacted with *Pcog-1* and FOZI-1 bound to *Plim-6* (Fig. 1B). Johnston et al. (2006) previously reported that DIE-1 and FOZI-1 negatively regulate the expression of *cog-1* and *lim-6*, respectively. Together with our Y1H results, this suggests that DIE-1 and FOZI-1 may act as direct transcriptional repressors of *cog-1* and *lim-6*.

We also identified three interologs or regulogs, i.e., evolutionarily conserved PDIs (Walhout et al. 2000a; Yu et al. 2004): VAB-3 bound to *Pdac-1*, TTX-1 bound to *Punc-30*, and UNC-30 bound its own promoter (Fig. 1B). During eye development in *Drosophila*, *eyeless*, the ortholog of VAB-3, activates the expression of *dac*, the ortholog of *dac-1* (Page et al. 2001). In mouse, it has been shown that Pitx1 and Pitx2, orthologs of UNC-30, and Otx1, an ortholog of TTX-1, positively regulate the *Pitx1* promoter (Goodyer et al. 2003). Together, this suggests that VAB-3 is a direct transcriptional activator of *dac-1* and that TTX-1 and UNC-30 can both directly activate the expression of *unc-30*.

Functional properties of the core neuronal PDI network

We characterized the functional properties of the network by available annotations of the interactor TFs. Since we chose a set of neuronal TF-encoding genes as targets in Y1H assays, we hypothesized that the core neuronal PDI network may be enriched for neuronally expressed interactor TFs, but not for TFs that are expressed in other tissues. In WormBase version WS153 (http://www.wormbase.org), expression pattern information was available for 293 of the 940 predicted *C. elegans* TFs (V. Vermeirssen, B. Deplancke, M.I. Barrasa, J.S. Reece-Hoyes, H.E. Arda, C.A. Grove, N.J. Martinez, R. Sequerra, L. Doucette-Stamm, M.R. Brent, et al., in prep.). WS153 contained expression pattern information for 66% of the interactor TFs in the network. Of all 293 TFs for which expression patterns are available, 63% had a neuronal expression pattern and 73% were expressed in multiple tissues, suggesting an extensive "re-use" of TFs in different biological processes. Despite this fact, we did observe a significant enrichment for TFs expressed in neurons in the core neuronal PDI network (76%, $P < 0.05$), but not for TFs expressed in the digestive tract, epithelial system, reproductive system, muscle, or excretory system in *C. elegans* (Table 1).

The interactor TFs in the core neuronal PDI network were significantly enriched for the homeodomain DNA-binding domain (28% in the core neuronal PDI network versus 10% in all predicted TFs, $P < 0.001$; Fig. 2). This was mainly due to the over-representation of the paired homeodomain ($P < 0.001$; Fig. 2, inset). In addition, the interactor TFs in the core neuronal PDI network were significantly depleted for the nuclear hormone receptor DNA-binding domain ($P < 0.001$; Fig. 2). Nuclear hormone receptors often bind DNA as dimers (Fuller 1991). Our current version of the Y1H system is not configured to detect PDIs with heterodimers, which could explain the smaller proportion of nuclear hormone receptors in the network. However, as discussed below, this smaller proportion could potentially be biologically relevant.

To further functionally characterize the interactor TFs, we inspected their Biological Process terms in the Gene Ontology (GO) database (Ashburner et al. 2000) (Supplemental Table S8). Compared with all predicted TFs in *C. elegans*, the interactor TFs in the core neuronal PDI network were significantly enriched for the GO terms "development" ($P < 0.01$) and "response to stimulus" ($P < 0.001$), both of which could be expected due to the selection of the target genes. Surprisingly, we also found an enrichment for the GO term "reproduction" ($P < 0.01$).
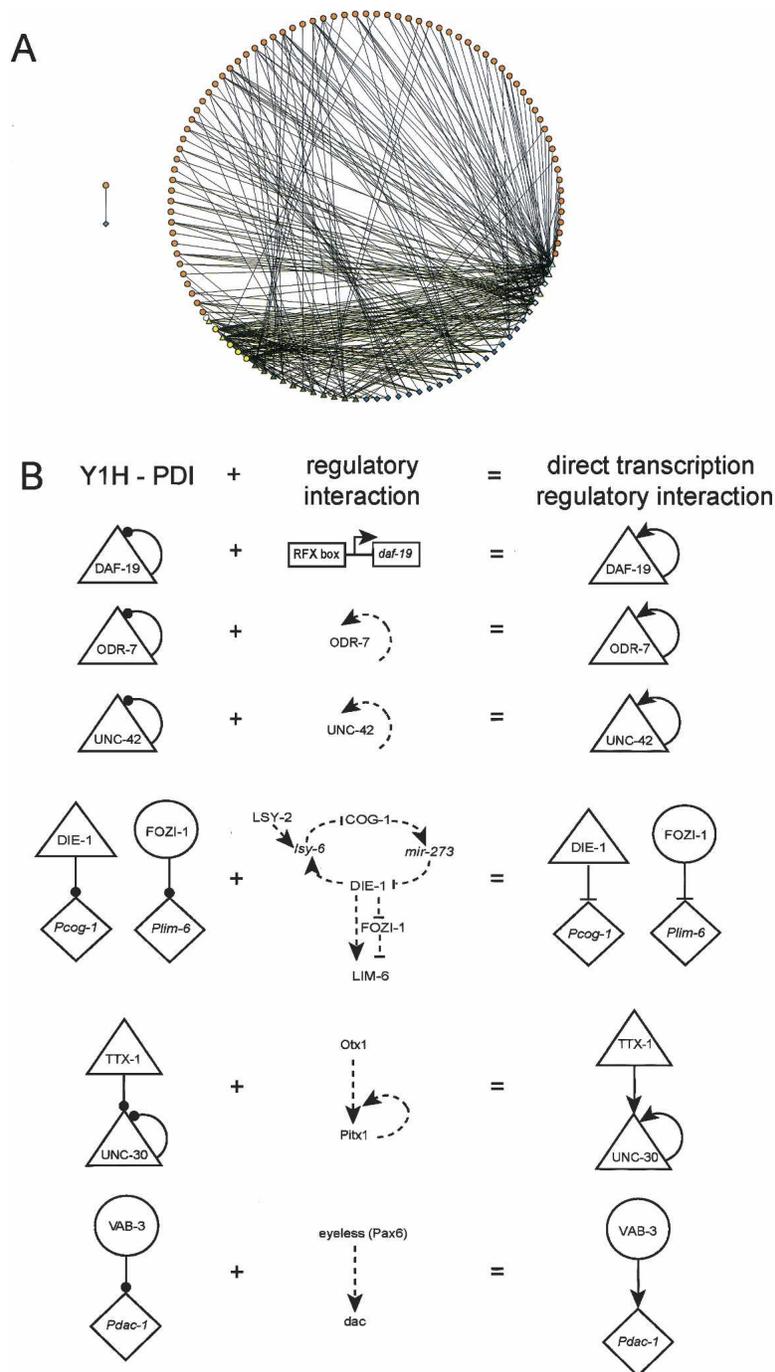
Figure 1. A *C. elegans* core neuronal PDI network. (*A*) The core neuronal PDI network visualized using Cytoscape v2.3 (for a larger depiction, see Supplemental Fig. S4). (*B*) From PDIs to transcription regulatory interactions by data integration. (*Top*) Three autoregulatory PDIs confirmed proposed regulatory interactions. DAF-19 was proposed to regulate its own expression because the *daf-19* promoter contains a predicted DAF-19 binding site (RFX box). (*Middle*) Two new PDIs may be involved in the determination of bilateral asymmetry in neurons. (*Bottom*) Two putative interologs/regulogs. Dotted lines with arrow (activation) or blunt arrow (repression), regulation; solid lines with dotted arrow, physical interaction.

that associate with GO terms that relate to both organism development and function.

The core neuronal PDI network contains both interactor and promoter hubs

To explore the architecture of the core neuronal PDI network, we first examined the degree distribution of both the promoter and the interactor nodes. As expected, the outgoing connectivity of the core neuronal PDI network followed a power law (Fig. 3A) (Barabasi and Albert 1999). This can be attributed to the presence of "interactor hubs," i.e., TFs that bind to a disproportionately large number of promoters. The incoming connectivity could be fit by a power law with saturation (Fig. 3B; Supplemental Materials). Promoters were on average bound by 7.4 interactors (i.e., the mean), whereas the median was four interactors (Supplemental Table S5). The difference in mean and median and the incoming degree distribution indicates the presence of "promoter hubs," i.e., promoters that are bound by a disproportionately large number of TFs. Indeed, the promoters *Pnhr-79*, *Pcog-1*, *Pdaf-3*, and *Punc-30* are bound by 22, 25, 27, and 36 interactors, respectively. In contrast to interactor hubs, none of these promoter hubs are essential for viability (Kamath et al. 2003; Sonnichsen et al. 2005).

The digestive tract PDI network, which contains target genes expressed in the *C. elegans* pharynx, intestine, and/or rectum (Deplancke et al. 2006a), contains a number of interactor hubs that bind to target genes that are expressed in both the pharynx and the intestine, even though these organs are derived from distinct germ layers. Therefore, we hypothesized that these interactor hubs may function as global regulators of gene expression i.e., they control the expression of many genes expressed in many different tissues. Eighty three percent of the interactor hubs (here defined as the 5% most highly connected interactors) in the core neuronal PDI network were also retrieved as interactor hubs in the digestive-tract PDI network, whereas only 47% of the less well-connected interactors overlap (Fig. 3C). This suggests that *C. elegans* genes that are expressed in different tissues may indeed be under the control of the same global regulators and that less well-connected interactors may function as specifiers in each tissue. One interactor hub in the core neuronal PDI network, C32D5.1, was not retrieved in the digestive-tract net-

Taken together, we found that the core neuronal PDI network is enriched for interactor TFs that are neuronally expressed, that possess a paired homeodomain DNA-binding domain, and

Table 1. The core neuronal PDI netowork is enriched for neuronally expressed interactor TFs

Tissue	TFs with expression patterns in core neuronal PDI network	Total TFs with expression patterns	P-value for enrichment
Total	55	293	
Neurons	42	186	0.0186[a]
Digestive tract	29	173	0.9071
Epithelial system	30	144	0.2300
Muscle	18	87	0.3467
Reproductive system	22	118	0.5762
Excretory system	5	40	0.9103

[a]Significant enrichment.

work (Deplancke et al. 2006a). By directed Y1H assays, we found that C32D5.1 can bind a similar proportion of digestive tract promoters as neuronal promoters (Supplemental Fig. S3), suggesting that it was missed in the digestive tract cDNA library screens and that C32D5.1 may also function as a global regulator. C32D5.1 does not have a predicted DNA-binding domain (Reece-Hoyes et al. 2005), but we confirmed that C32D5.1 can bind directly to its target genes by ChIP in yeast (Supplemental Fig. S3). Moreover, it does share 58% similarity and 38% identity with the predicted MADF DNA-binding domain of Y55F3BR.5 (Supplemental Fig. S3). Together, these observations suggest that C32D5.1 may indeed be a DNA-binding protein, potentially with a MADF-type DNA-binding domain.

Cross-regulation

Cross-regulation occurs when, within a system of interest (i.e., cell, tissue, or organ), many TFs regulate each other's expression. Previously, cross-regulation has been observed in both human and yeast regulatory networks (Borneman et al. 2006; Odom et al. 2006). In the core neuronal PDI network, we also found significant cross-regulation, as 16 of the 94 interactors corresponded to TFs that were included in the target gene set ($P < 0.001$). Interestingly, multiple target genes encode homeodomain TFs (41%), and the core neuronal PDI network is enriched for homeodomain interactor TFs (Fig. 2). Similarly, both the target genes and many interactor TFs are expressed in neurons (Table 1). However, the cross-regulation was not solely responsible for either the homeodomain or neuronal enrichment of the interactor TFs in the network (data not shown). Taken together, we find a significant level of cross-regulation in the core neuronal PDI network.

The core neuronal PDI network contains two TF modules

Metabolic and protein–protein interaction networks can often be decomposed into functional modules: groups of highly interconnected components that together carry out particular biological functions (Ravasz et al. 2002; Han et al. 2004; Yook et al. 2004; Gunsalus et al.

2005). To examine whether the core neuronal PDI network possesses a modular architecture, we analyzed the topological overlap coefficient (TOC) or mutual clustering coefficient for each pair of nodes (Ravasz et al. 2002; Goldberg and Roth 2003). This is a relative measure for the number of interaction partners shared by a pair of nodes, and ranges from 0 for node pairs that do not share any interacting partners to 1 for node pairs that share all interacting partners (Fig. 4A). Since PDI networks are directed networks, we calculated the TOC separately for each pair of promoters and each pair of interactors. The TOC for a pair of promoters defines the overlap of their interactors. The TOC for a pair of interactors defines the overlap of the promoters they interact with. We used two formulas, the meet/min and the geometric formula, both of which gave similar results (Fig. 4B; Supplemental Fig. S5) (Goldberg and Roth 2003). The resulting topological overlap matrix was clustered in order to place nodes with similar TOCs close to each other. The topological overlap matrix of promoters exhibited no modularity (data not shown). In the topological overlap matrix of interactors, however, we observed two clusters with similar TOCs, referred to as module 1 and module 2 (Fig. 4B; Supplemental Fig. S5). As a control, we performed the same analysis for randomized networks with the same single node characteristics and did not observe modularity (data not shown). By visualizing a topological overlap network (Fig. 4A), which linked interactors that had a TOC above a threshold value, we confirmed that the interactors could be separated into two modules (Fig. 4B; Supplemental Fig. S5).

Functional aspects of the core neuronal PDI network are reflected in the two TF modules

The two interactor modules could be traced back to different sets of promoter hubs (Fig. 4C). Module 1 consisted of interactors that bound at least two of the following promoter hubs: *Pdaf-3*, *Pcog-1*, or *Pnhr-79*. Module 2 contained interactors shared between the promoter hub *Punc-30* and nine other promoters, predominantly the promoter hub *Pceh-23*. When we further inspected these promoter hubs, we observed that *daf-3*, *cog-1*, and *nhr-79* (module 1) all encode TFs that are expressed and/or function in neurons and in other tissues as well, whereas *unc-30* and
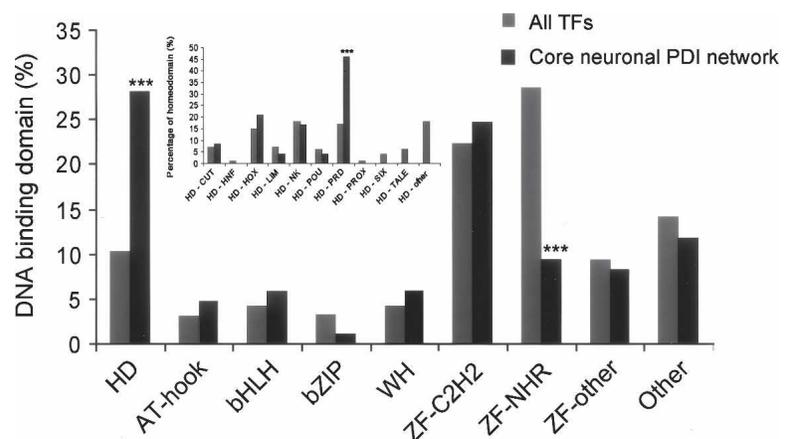
Figure 2. The core neuronal PDI network is enriched for paired homeodomains. The percentage of the different DNA-binding domains in TFs retrieved in the core neuronal PDI network versus all *C. elegans* TFs. The network is significantly enriched for TFs that possess a homeodomain (HD, $P < 0.001$), particularly paired homeodomains (HD-PRD, *inset*, $P < 0.001$), and is significantly depleted for nuclear hormone receptors (ZF-NHR, $P < 0.001$).
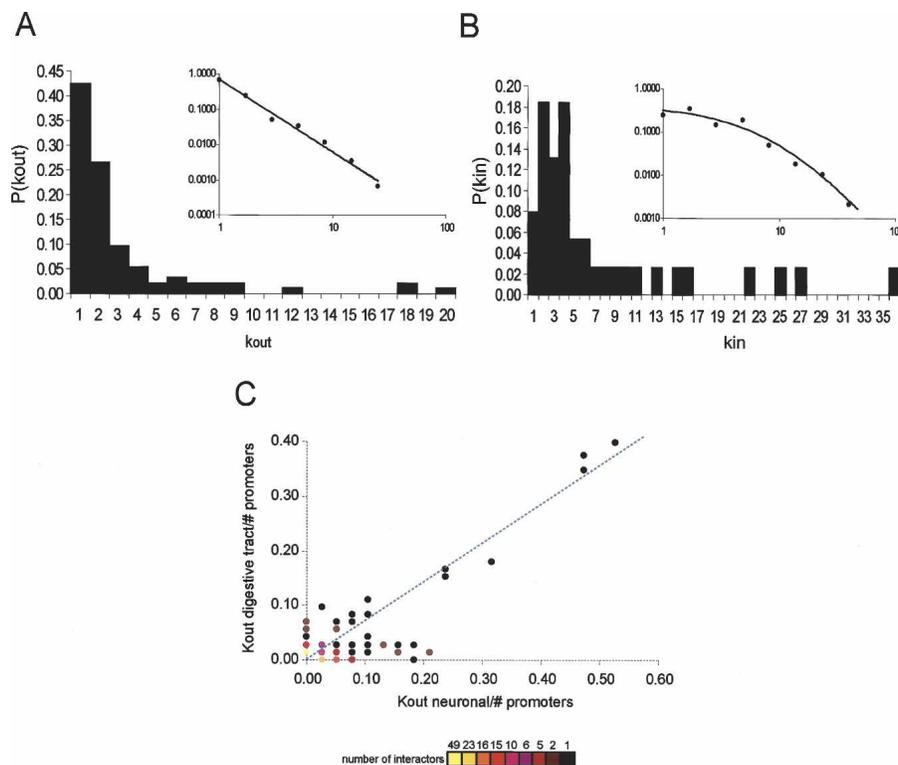
Figure 3. Architecture of the *C. elegans* core neuronal PDI network. (*A*) Out-degree distribution. Logarithmic binning was applied to fit a power law to $P(k_{out})$: $P(k_{out}) = 0.69k^{-2.05}$ ($R^2 = 0.99$; *inset*). (*B*) In-degree distribution. Logarithmic binning was applied to fit a power law with saturation to $P(k_{in}) = 1093.2*(10.83 + k)^{-3.30}$ ($R^2 = 0.97$; *inset*). (*C*) Interactor hubs in the core neuronal PDI network have a similar out-degree in the digestive tract PDI network (diagonal). Interactors specific for each network are localized on the axes. The colors indicate the number of interactors per feature.

half of the target genes in module 2 are exclusively expressed in neurons and function as neuronal cell-fate determinants, this suggests that the TFs these genes interact with may be specifically expressed in neurons as well.

Paired homeodomain TFs are expressed in neurons in both *C. elegans* and mouse

Since the core neuronal PDI network and module 2 were both enriched for interactor TFs that are expressed in neurons and that possess a (paired) homeodomain DNA-binding domain, we hypothesized that (paired) homeodomain-containing TFs are likely to be neuronally expressed. To test this, we analyzed the available expression patterns of all predicted *C. elegans* TFs in WS153. Indeed, we found a positive association between neuronal expression and homeodomains both for TFs that are expressed in neurons as well as in other tissues ($P < 0.01$) and for TFs that are exclusively neuronal ($P < 0.05$; Fig. 5A). Homeodomain TFs that are exclusively expressed in neurons tend to specifically possess a paired homeodomain ($P < 0.01$). We also found a negative association between neuronal expression and nuclear hormone receptors ($P < 0.001$), suggesting the depletion of this family in the core neuronal PDI network may be biologically relevant.

ceh-23 (module 2) are predominantly expressed in neurons and function in the determination of neuronal cell fate. In addition, half of the other target genes of module 2 are exclusively expressed in neurons (Supplemental Table S6). In total, 29% of the interactors and 21% of the PDIs were present in the two modules. The interactors in both modules have a low out-degree (Fig. 4C; Supplemental Table S6), suggesting that these TFs are specifiers (Deplancke et al. 2006a). Interactors with a higher average degree, including the global regulators, connect the two modules (Supplemental Table S6).

Next, we asked whether the functional properties of the core neuronal PDI network were reflected in the modules. As expected, both modules contained many neuronally expressed TFs (Supplemental Table S6). However, we found several functional differences between the two modules. Only module 1 was enriched for the GO terms "reproduction" and "response to stimulus" ($P < 0.001$, Supplemental Table S8). Since the numbers are small, we also investigated the available mutant and RNAi phenotypes as documented in WormBase (http://www.wormbase.org) and found the same results (Fig. 4C; Supplemental Table S6). These GO enrichments correlate with the functional annotation of the target genes: *cog-1* is involved in vulva formation and *cog-1* mutants are egg-laying defective and sterile (Page et al. 2001), whereas *daf-3* functions in the TGF-β pathway, which transports sensory cues of the environment to the egg-laying and dauer-formation circuits (Patterson et al. 1997). Module 2 was significantly enriched for homeodomain TFs ($P < 0.001$) and especially paired homeodomains ($P < 0.01$). Since

Next, we examined whether the positive association between paired homeodomains and neuronal expression is evolutionarily conserved. We analyzed murine tissue-specific TF gene expression profiles available in SymAtlas (Su et al. 2004). We also observed a significant enrichment for murine homeodomain TFs that are exclusively expressed in neurons ($P < 0.001$; Fig. 5B). Moreover, we noted once more that these TFs tend to possess a paired homeodomain ($P < 0.05$), which complements the observations made for the *C. elegans* core neuronal PDI network. To a lesser extent, homeodomains were also enriched in murine endocrine and reproductive systems. However, paired homeodomains were not responsible for the enrichment in these tissues (Fig. 5B). In two other datasets of TF gene expression in mouse and human, we also noticed that neuronal tissues specifically express homeodomain TFs (Gray et al. 2004; Kong et al. 2006; data not shown). We analyzed the GO terms of murine-paired homeodomain TFs and noticed an enrichment for "central nervous system development," "brain development," "eye development," "neurogenesis," "neuron differentiation," and "neuron migration" (data not shown). We conclude that homeodomain TFs in general, and paired homeodomain TFs in particular, likely regulate neuronal gene expression during development in both *C. elegans* and in mouse.

Finally, we noticed additional correlations between TF families and expression patterns (Fig. 5). In the future, such correlations may help to decipher the transcription regulatory networks of particular tissues, organs, and cell types.
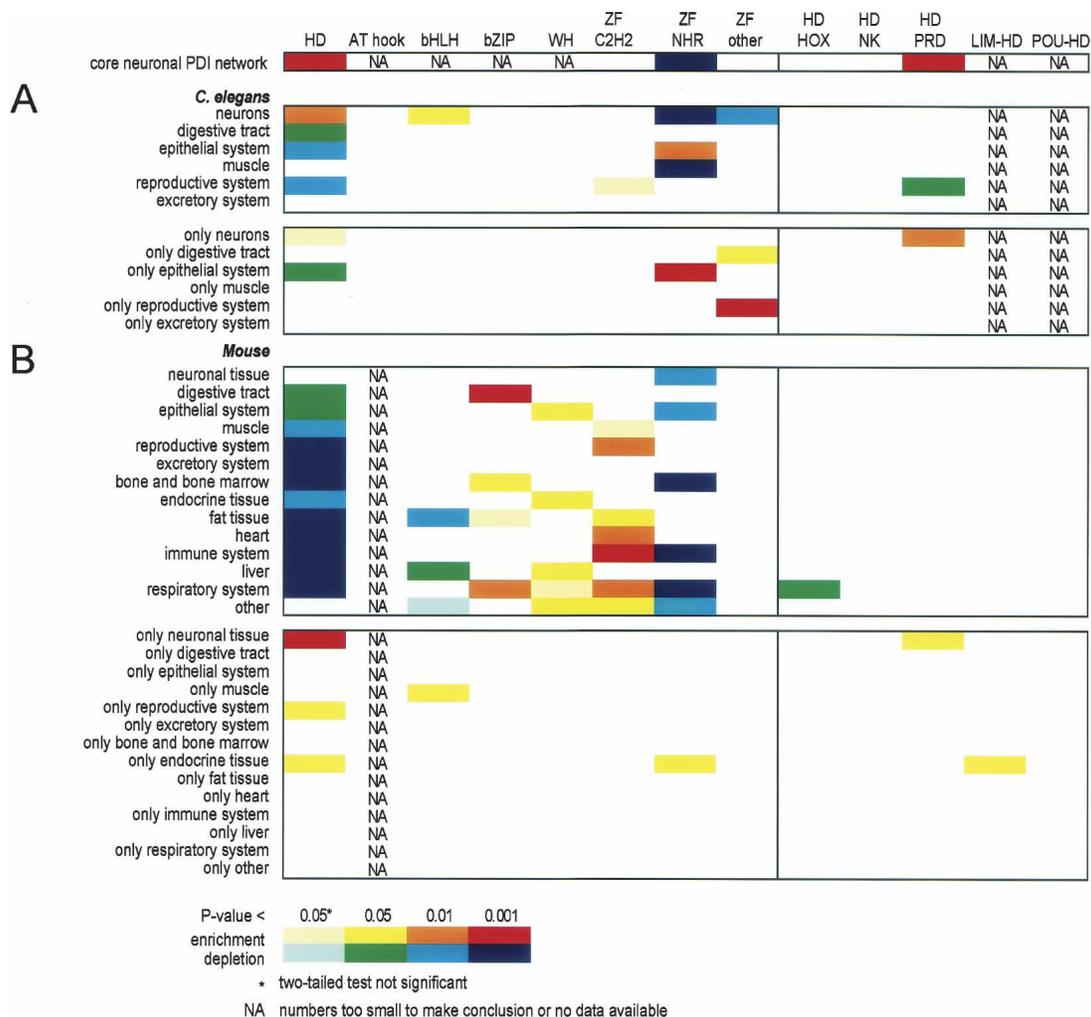
Figure 4. The core neuronal PDI network contains two specifier TF modules. (*A*) From a PDI network to a topological overlap network. (*B*) Topological overlap matrix for the interactors using the geometric formula and the resulting topological overlap network for interactor pairs with a topological coefficient ≥0.2. (Pink) Interactors in module 1 that connect to the promoter hubs *Pdaf-3*, *Pcog-1*, and *Pnhr-79*; (green) interactors in module 2 that bind the promoter hub *Punc-30*, in addition to *Pceh-23*, *Pnhr-83*, *PT22H9.4*, *Pnhr-41*, *Pttx-1*, *Pdaf-19*, *Phlh-2*, *Pnhr-6*, or *Plin-32*. (See Supplemental Fig. 5 for the modules obtained with the meet/min formula.) (*C*) PDI network corresponding to the interactors indicated in *B*. (Red) Paired homeodomain; (orange) other homeodomain, (R) reproduction; (B) response to stimulus (behavior). The average outgoing degree is indicated.

Discussion

Gene-centered PDI networks

In this study, we present a gene-centered PDI network of neuronal TF-encoding genes in the nematode *C. elegans*. This network contains 282 high-confidence PDIs, between 38 promoters and 94 interactors. Most of the PDIs are novel and most TFs retrieved were heretofore uncharacterized. This demonstrates that gene-centered approaches rapidly expand our knowledge about PDIs and help annotate both individual TFs and TF families (see below).

Several observations demonstrate that the Gateway-compatible Y1H system yields high-quality PDIs. First, we show that a stringent and standardized scoring system can be used to extract high-confidence Y1H interactions. Second, we show that networks derived from sets of target genes expressed in a particular tissue are enriched for interactors that are also expressed in that tissue (this study; Deplancke et al. 2006a). Third, by integrating PDI data with regulatory information, we convert PDIs into transcription regulatory interactions. Fourth, we find several interologs. Fifth, many PDIs can be connected to previously reported observations. For instance, we find interactions between factors and target genes that are expressed in the same cell(s), e.g., *Pnhr-38*/CEH-14 in AFD sensory neurons (Miyabayashi et al. 1999; Cassata et al. 2006) and *Punc-30*/ALR-1 in GABAergic neurons (Melkman and Sengupta 2005). In addition, we find a putative regulatory cascade that may be involved in dauer formation: DAF-16 binds the *daf-3* promoter, DAF-3 binds the *daf-19* promoter, and DAF-19 regulates its own expression.

Network architecture

Cellular networks are characterized by a scale-free connectivity distribution due to the presence of highly connected nodes, or hubs (Barabasi and Oltvai 2004). Our observation that *C. elegans* interactor hubs connect to genes expressed in different tissues and cell types suggest that such global regulators function throughout the animal. In agreement with this, we previously found that global regulators tend to be broadly expressed, to be essential for viability, and to be toxic when overexpressed (Deplancke et al. 2006a). We find that interactors with a low out-degree tend to be

Figure 5. Homeodomain TFs are neuronally expressed both in *C. elegans* and in mouse. (*A*) Association analysis between TF family and expression patterns in *C. elegans*. (*B*) Association analysis between TF family and expression profiles in mouse. *P*-values were calculated by a two-tailed test, as well as by a right- and left-tailed test for enrichment and depletion, respectively. We only show the latter values, and indicate with an asterisk when there was a slight enrichment/depletion for which the two-tailed test was not significant.

more specific for either the neuronal or the digestive tract PDI network. Together, the finding of global and specifier regulators supports a model of a layered organization of TF function in *C. elegans* transcription regulatory networks (Deplancke et al. 2006a).

The in-degree distribution of transcription regulatory networks in yeast and bacteria has been reported to decay exponentially, which is similar to the degree distribution of random networks, and suggests that there are no clear promoter hubs (Thieffry et al. 1998; Guelzim et al. 2002). However, highly connected promoters have been described in yeast regulatory networks (Yu et al. 2004; Borneman et al. 2006). The incoming degree distribution of the core neuronal PDI network was best fit by a power law with saturation. Although the biological significance of this fit is at present unclear, it may result from the gene duplication driven growth and evolution of (biological) networks (Albert and Barabasi 2000; Vazquez et al. 2003). The incoming degree distribution points to the presence of promoter hubs. Indeed, we do find several promoter hubs such as *Punc-30* that bind 36 different TFs. Such promoter hubs may be specific for core

transcription regulatory networks, because the promoters of TF-encoding genes have been proposed to be subject to more complex regulation than other genes (Nelson et al. 2004; Woolfe et al. 2005). Future studies with non-TF genes are required to determine whether the in- and out-degree distribution of the core neuronal PDI network are a reflection of the complete *C. elegans* transcription regulatory network. Previously, it has been suggested that highly regulated TFs may function as master regulators of development (Borneman et al. 2006). This notion is supported by the fact that UNC-30, the most highly connected promoter hub in our network, functions as a master regulator for the terminal differentiation of type-D GABAergic motor neurons (Jin et al. 1994).

TF modularity in PDI networks

Through a combination of PDI mapping, network analysis, TF family annotation, and gene expression and ontology analysis, we demonstrate for the first time that a metazoan PDI network is organized into TF modules that relate to specific functionalities. Previously, it has been shown that regulatory networks from uni-

cellular systems such as bacteria and yeast possess a modular architecture (Babu et al. 2004; Resendis-Antonio et al. 2005). Bacterial PDI network modularity was found by clustering the shortest path length between any pair of genes in an undirected manner (Resendis-Antonio et al. 2005). Since PDI networks contain directed interactions, two types of modules can potentially occur: gene modules that contain genes that share interacting TFs or TF modules that contain TFs that share target genes. Gene modules have previously been identified by expression-profiling correlation of TFs and their target genes (Tavazoie et al. 1999; Segal et al. 2003; Ghazalpour et al. 2006) and by integrating such data with physical interaction data obtained from TF-centered approaches (Bar-Joseph et al. 2003; Beyer et al. 2006). We did not observe any target-gene modularity in the core neuronal PDI network. This is likely because we only used promoters of neuronal genes and because most promoters bind a combination of global and specifier TFs, most of which bind only one or two promoters (Fig. 3A). We expect that target modularity will be uncovered in larger *C. elegans* PDI networks that contain genes involved in different biological processes and/or expressed in different tissues.

We did find two TF modules in the core neuronal PDI network. In metazoan PDI networks, such TF modularity is uniquely revealed by a gene-centered approach, which enables the identification of numerous TFs (94 interactors in this study) in a standardized manner. Two TF modules consist of specifier TFs and are connected to each other by several other TFs, including all of the putative global regulators. Moreover, we find that the two modules associate with different functional aspects of the total network. One of the most striking findings is that module 2 is enriched for paired homeodomain TFs and contains target genes that are predominantly expressed in neurons (which is in contrast to the targets of module 1 that are expressed more broadly). This correlation suggests that paired homeodomain TFs specifically regulate neuronal gene expression. Indeed, we find that paired homeodomain TFs tend to be exclusively expressed in neurons in both *C. elegans* and mouse. Several homeodomain genes are known to play a role in the development of the central nervous system in *Drosophila* and vertebrates (Kammermeier and Reichert 2001; Akin and Nazarali 2005). Recently, an over-representation of homeodomain binding sites has been detected in the promoters of odorant receptors in mouse (Michaloski et al. 2006), and some of these were shown to be required for normal odorant receptor expression (Rothman et al. 2005). This further confirms that homeodomain TFs function in sensory neurons. Several individual *C. elegans* and murine-paired homeodomain TFs are known to function in neurons (Miller et al. 1992; Jin et al. 1994; Baran et al. 1999; Pujol et al. 2000; Altun-Gultekin et al. 2001; Boyl et al. 2001; Satterlee et al. 2001; Lanjuin et al. 2003; Branicky and Hekimi 2005; Melkman and Sengupta 2005; Tucker et al. 2005; Friocourt et al. 2006). Our data suggest that uncharacterized (paired) homeodomain TFs may also function to regulate neuronal gene expression and function.

Multiple paired homeodomain TFs in module 2 bind the same target promoters, and may perhaps interact with the same site within these promoters. Since we found that paired homeodomain TFs tend to be specifically expressed in neurons, this suggests that several members of this TF family may bind the same neuronal gene promoter in vivo, for instance, under different conditions during development or function of the animal. Future studies, for instance, using ChIP, will be required to determine whether all or some of these TFs all interact with their Y1H targets in vivo. In addition, it will be important to determine which

of the interacting TFs actually affect target gene expression in vivo and where and when in the (developing) animal these effects occur.

Taken together, by integrating physical interactions with regulatory events, TF families, expression patterns and profiles, and functional annotations, we show that metazoan PDI networks have a modular architecture that relates to network functionality. Such modularity will provide a powerful tool to understand how networks relate to biology and, potentially, to annotate gene function in complex metazoan organisms.

Methods

Generation of Y1H promoter bait strains

Detailed Y1H protocols are described elsewhere (Deplancke et al. 2006b). Promoters for 47 neuronal transcription regulatory genes were selected as DNA baits (Supplemental Table S1). For *daf-16* and *daf-12*, two promoters were selected, based on different variants of each gene. For *ceh-37*, two promoter baits were created, one upstream of the ATG and one upstream of the 5′-UTR, since this is >10 kb upstream of the first exon. Twenty promoter Entry clones were retrieved from the promoterome (Dupuy et al. 2004). We attempted to clone 30 promoters ab initio (i.e., by PCR from genomic DNA using the translational start reported in WS153), and cloned all except *Pdaf-12a* and *Pnhr-36* (Supplemental Fig. S1). With the exception of *Pfozi*-1, all promoters were transferred into the Y1H destination vectors pMW#2 and pMW#3 and integrated into the genome of *Saccharomyces cerevisae* YM4271. Promoters that exhibit high self-activation for the *lacZ* reporter could only be analyzed for activation of the *HIS3* reporter (Supplemental Table S2). *Pdaf-12b::lacZ* could not be integrated into the yeast genome, and thus, only the *HIS3* reporter could be used for this promoter.

Y1H screens were performed with individual DNA bait strains versus both AD-wrmcDNA (Walhout et al. 2000b) and AD-TF (Deplancke et al. 2004) prey libraries (Supplemental Fig. S1). Only double positives were considered, except when *lacZ* was absent or highly self-active. All PDIs were retested by PCR/gap repair. PCR products corresponding to preys that retested were sequenced by Agencourt Bioscience Corporation. In total, 805 Interaction Sequence Tags were obtained (i.e., 5′ tag sequences of the interactor preys) (Walhout et al. 2000a). The PDI data for *Pdaf-3*, *Pdaf-19*, *Pdie-1*, and *PT22H9.4* were complemented by results obtained previously (Deplancke et al. 2006a).

Y1H mating assays against an AD-TF array (V. Vermeirssen, B. Deplancke, M.I. Barrasa, J.S. Reece-Hoyes, H.E. Arda, C.A. Grove, N.J. Martinez, R. Sequerra, L. Doucette-Stamm, M.R. Brent, et al., in prep.) were performed with 16 bait strains (Supplemental Table S2).

Finally, all available interactors were transformed into each of the promoter strains. In addition to verifying PDIs, this enables the identification of additional PDIs (Supplemental Fig. S1). Preys found by Y1H screens and mating experiments were used, as well as TFs corresponding to the Y1H promoter baits (150 preys were used in total; Supplemental Table S3). The following preys could not be examined because an ORF clone was not available: HMG-11, LIN-49, NHR-36, NHR-45, NHR-83, R06C1.6, UNC-86, Y59E9AL.2/3, and ZK287.6. Ninety-nine percent of the 7050 transformations were successful.

We performed these different Y1H assays because screens alone are not saturating. Many TFs are likely under-represented because the cDNA library is not normalized, and may therefore be difficult to retrieve when a few million colonies are screened. We find that we obtain the highest coverage by screening both libraries. Some TFs cannot be detected when full-length proteins

are used in Y1H. However, since the cDNA library contains many incomplete ORFs, these may be retrieved from this library. In addition, the cDNA library enables the identification of novel putative TFs.

C. elegans TF expression patterns

Only interactors possessing a predicted DNA-binding domain were considered (i.e., novel putative TFs were excluded). Expression pattern information for the 940 predicted *C. elegans* TFs in wTF2.1 (V. Vermeirssen, B. Deplancke, M.I. Barrasa, J.S. Reece-Hoyes, H.E. Arda, C.A. Grove, N.J. Martinez, R. Sequerra, L. Doucette-Stamm, M.R. Brent, et al., in prep.) was retrieved using WormMart in WormBase WS153. An expression-pattern code was attributed to each TF as follows: (E) embryonic, (G) germline, (H) epithelial tissues except for neuronal support cells, (I) intestine, (M) muscle, (N) neuron, (O) other or unidentified cells, (P) pharynx, (R) somatic reproductive tissue, (S) neuronal support cells (socket/sheath/glial cells), and (X) excretory system. These can be grouped to describe expression in the following systems: (HS) epithelial system, (PI) digestive tract, and (GR) reproductive system. In the reproductive and digestive tract systems, only expression in the structural parts (i.e., musculature, glands, epithelium) was annotated, while expression in the neurons within those systems was considered N. Some examples: pharyngeal neurons (N), vulval muscle (MR), rectal epithelium (HI), amphids or phasmids (NS), male sensory rays (NS), broad (EGHIMNOPRSX), and all somatic cells (EHIMNOPRSX). Embryonic expression was not included in subsequent analyses. Expression pattern information was available for 293 (N) of the 940 predicted TFs in wTF2.1. Information was available in WS153 for 55 (*n*) of the 83 interactor TFs in the core neuronal PDI network. Enrichment for expression in a certain tissue was calculated using a hypergeometric distribution. For a set of *n* (55) interactor TFs, of which *k* are annotated with an expression pattern in a certain tissue that exists in *K* of the *N* (293) *C. elegans* TFs, the hypergeometric *P*-value is given by:

$$P(X \geq k) = \sum_{i=k}^{n} \frac{\binom{K}{i}\binom{N-K}{n-i}}{\binom{N}{n}} .$$

C. elegans DNA-binding domains

DNA-binding domains for the 940 predicted *C. elegans* TFs were extracted from wTF2.1 (V. Vermeirssen, B. Deplancke, M.I. Barrasa, J.S. Reece-Hoyes, H.E. Arda, C.A. Grove, N.J. Martinez, R. Sequerra, L. Doucette-Stamm, M.R. Brent, et al., in prep.). Only DNA-binding domain classes with 30 or more members were considered (homeodomain, AT hook, basic helix-loop-helix, bZIP, winged helix, zinc finger, and others). Zinc fingers were divided into C2H2 zinc fingers, nuclear hormone receptors, and other zinc fingers. By a χ^2 test we first analyzed whether the nodes in the network represent a random sample of wTF2.1. Any class with less than five expected members in the χ^2 test was added to the class of "other DNA-binding domains." The enrichment or depletion for a specific DNA-binding domain in the network was determined by a hypergeometric distribution ($N = 962$ DNA binding domains—some TFs possess multiple DNA-binding domains). A similar approach was followed to analyze the DNA-binding domain distribution within the homeodomain class.

Gene Ontology analysis

The enrichment of Biological Process Gene Ontology terms for the interactor TFs in the network or modules was calculated by a Fisher test using only genes for which such terms were available in WormBase (WS164). They were available for 83% of the interactor TFs in the network and for 82% and 100% of the TFs in

modules 1 and 2, respectively. Biological Process Gene Ontology terms were available for 76% of all 940 predicted *C. elegans* TFs. This was used as background population in the statistical analysis. We calculated the enrichment in the complete network for the following terms: reproduction, development, physiological process, growth, cellular process, regulation of biological process, and response to stimulus. In the modules, we only examined the three GO terms that were significantly enriched in the entire network (reproduction, development, and response to stimulus). Nominal *P*-values are included in the main text. We also corrected for multiple hypothesis testing by applying a Bonferroni correction (Supplemental Table S8). Finally, we confirmed the statistical enrichments by an independent method, FuncAssociate, which is based on random sampling (Berriz et al. 2003; data not shown).

The enrichment of Biological Process Gene Ontology terms for the paired homeodomains in mouse was calculated by the DAVID functional annotation tool (http://david.abcc.ncifcrf.gov/) (Dennis et al. 2003).

Cross-regulation

The enrichment for cross-regulation, i.e., the retrieval of interactor TFs of which the promoters were also used as DNA baits, was calculated by a hypergeometric distribution (see above) using all 940 TFs (*N*) as the population. In total, 16 (*k*) of the 36 TF-encoding target genes (*K*) were retrieved as interactors of the 83 TFs (*n*) in the network.

C. elegans DNA–binding domain—Expression pattern association

WS153 was used to examine whether there is an association (Fisher test) between the presence of a particular DNA-binding domain in a TF and the expression of this TF in a particular *C. elegans* tissue. A total of 304 expression patterns were available for all DNA-binding domains (some TFs contain multiple DNA-binding domains). WS153 contained expression information for 69% of all homeodomains, 23% of all AT hooks, 51% of all basic helix-loop-helices, 25% of all bZIPs, 56% of all winged helices, 21% of all C2H2 zinc fingers, 20% of all nuclear hormone receptors, 35% of all other zinc fingers, and 32% of all other DNA-binding domains in wTF2.1. Within the homeodomain class, HOX, paired, and NK homeodomains were examined. The following tissues were considered: neurons, muscle, digestive tract, epithelial system, reproductive system, and excretory system.

Mouse DNA-binding domains

Mouse TFs were compiled by first downloading mouse TF predictions from DBD, a transcription factor database (version 1.2) (Kummerfeld and Teichmann 2006). Ensembl protein IDs were mapped to gene IDs using BioMart (http://www.ensembl.org/Multi/martview) and all predictions for a gene were merged. A total of 1305 mouse TFs that possessed a total of 1421 DNA-binding domains were obtained. To identify the specific types of homeodomain within the homeodomain class, the INTERPRO (version 12.1) identifiers IPR007104 for paired, IPR007107 for LIM, and IPR007103 for POU homeodomains were used. HOX and NK homeodomains were identified from the literature, as INTERPRO identifiers are not available for these (Luke et al. 2003; Akin and Nazarali 2005).

Additional Methods are available in Supplemental Materials.

Acknowledgments

References

Akin, Z.N. and Nazarali, A.J. 2005. Hox genes and their candidate downstream targets in the developing central nervous system. *Cell. Mol. Neurobiol.* **25:** 697–741.

Albert, R. and Barabasi, A.L. 2000. Topology of evolving networks: Local events and universality. *Phys. Rev. Lett.* **85:** 5234–5237.

Altun-Gultekin, Z., Andachi, Y., Tsalik, E.L., Pilgrim, D., Kohara, Y., and Hobert, O. 2001. A regulatory cascade of three homeobox genes, *ceh-10, ttx-3* and *ceh-23*, controls cell fate specification of a defined interneuron class in *C. elegans*. *Development* **128:** 1951–1969.

Ashburner, M., Ball, C.A., Blake, J.A., Botstein, D., Butler, H., Cherry, J.M., Davis, A.P., Dolinski, K., Dwight, S.S., Eppig, J.T., et al. 2000. Gene ontology: Tool for the unification of biology. The Gene Ontology Consortium. *Nat. Genet.* **25:** 25–29.

Babu, M.M., Luscombe, N.M., Aravind, L., Gerstein, M., and Teichmann, S.A. 2004. Structure and evolution of transcriptional regulatory networks. *Curr. Opin. Struct. Biol.* **14:** 283–291.

Bar-Joseph, Z., Gerber, G.K., Lee, T.I., Rinaldi, N.J., Yoo, J.Y., Robert, F., Gordon, D.B., Fraenkel, E., Jaakkola, T.S., Young, R.A., et al. 2003. Computational discovery of gene modules and regulatory networks. *Nat. Biotechnol.* **21:** 1337–1342.

Barabasi, A.L. and Albert, R. 1999. Emergence of scaling in random networks. *Science* **286:** 509–512.

Barabasi, A.L. and Oltvai, Z.N. 2004. Network biology: Understanding the cell's functional organization. *Nat. Rev. Genet.* **5:** 101–113.

Baran, R., Aronoff, R., and Garriga, G. 1999. The *C. elegans* homeodomain gene *unc-42* regulates chemosensory and glutamate receptor expression. *Development* **126:** 2241–2251.

Barrasa, M.I., Vaglio, P., Cavasino, F., Jacotot, L., and Walhout, A.J.M. 2007. EDGEdb: A transcription factor-DNA interaction database for the analysis of *C. elegans* differential gene expression. *BMC Genomics* **8:** 21.

Berriz, G.F., King, O.D., Bryant, B., Sander, C., and Roth, F.P. 2003. Characterizing gene sets with FuncAssociate. *Bioinformatics* **19:** 2502–2504.

Beyer, A., Workman, C., Hollunder, J., Radke, D., Moller, U., Wilhelm, T., and Ideker, T. 2006. Integrated assessment and prediction of transcription factor binding. *PLoS Comput. Biol.* **2:** e70.

Blais, A. and Dynlacht, B.D. 2005. Constructing transcriptional regulatory networks. *Genes & Dev.* **19:** 1499–1511.

Borneman, A.R., Leigh-Bell, J.A., Yu, H., Bertone, P., Gerstein, M., and Snyder, M. 2006. Target hub proteins serve as master regulators of development in yeast. *Genes & Dev.* **20:** 435–448.

Boyl, P.P., Signore, M., Annino, A., Barbera, J.P., Acampora, D., and Simeone, A. 2001. Otx genes in the development and evolution of the vertebrate brain. *Int. J. Dev. Neurosci.* **19:** 353–363.

Branicky, R. and Hekimi, S. 2005. Specification of muscle neurotransmitter sensitivity by a paired-like homeodomain protein in *Caenorhabditis elegans*. *Development* **132:** 4999–5009.

Cassata, G., Kagoshima, H., Andachi, Y., Kohara, Y., Durrenberger, M.B., Hall, D.H., and Burglin, T.R. 2006. The LIM homeobox gene *ceh-14* confers thermosensory function to the AFD neurons in *Caenorhabditis elegans*. *Neuron* **25:** 587–597.

Davidson, E. and Levine, M. 2005. Gene regulatory networks. *Proc. Natl. Acad. Sci.* **102:** 4935.

Davidson, E.H., McClay, D.R., and Hood, L. 2003. Regulatory gene networks and the properties of the developmental process. *Proc. Natl. Acad. Sci.* **100:** 1475–1480.

Dennis Jr., G., Sherman, B.T., Hosack, D.A., Yang, J., Gao, W., Lane, H.C., and Lempicki, R.A. 2003. DAVID: Database for annotation, visualization, and integrated discovery. *Genome Biol.* **4:** 3.

Deplancke, B., Dupuy, D., Vidal, M., and Walhout, A.J.M. 2004. A Gateway-compatible yeast one-hybrid system. *Genome Res.* **14:** 2093–2101.

Deplancke, B., Mukhopadhyay, A., Ao, W., Elewa, A.M., Grove, C.A., Martinez, N.J., Sequerra, R., Doucette-Stam, L., Reece-Hoyes, J.S., Hope, I.A., et al. 2006a. A gene-centered *C. elegans* protein–DNA interaction network. *Cell* **125:** 1193–1205.

Deplancke, B., Vermeirssen, V., Arda, H.E., Martinez, N.J., and Walhout, A.J.M. 2006b. Gateway-compatible yeast one-hybrid screens. *CSH*

Protocols doi:10.1101/pdb.prot4590.

Dupuy, D., Li, Q., Deplancke, B., Boxem, M., Hao, T., Lamesch, P., Sequerra, R., Bosak, S., Doucette-Stam, L., Hope, I.A., et al. 2004. A first version of the *Caenorhabditis elegans* promoterome. *Genome Res.* **14:** 2169–2175.

Elnitski, L., Jin, V.X., Farnham, P.J., and Jones, S.J.M. 2006. Locating mammalian transcription factor binding sites: A survey of computational and experimental techniques. *Genome Res.* **16:** 1455–1464.

Friocourt, G., Poirier, K., Rakic, S., Parnavelas, J.G., and Chelly, J. 2006. The role of ARX in cortical development. *Eur. J. Neurosci.* **23:** 869–876.

Fuller, P.J. 1991. The steroid receptor superfamily: Mechanisms of diversity. *FASEB J.* **5:** 3092–3099.

Ghazalpour, A., Doss, S., Zhang, B., Wang, S., Plaisier, C., Castellanos, R., Brozell, A., Schadt, E.E., Drake, T.A., Lusis, A.J., et al. 2006. Integrating genetic and network analysis to characterize genes related to mouse weight. *PLoS Genet.* **2:** e130.

Goldberg, D.S. and Roth, F.P. 2003. Assessing experimentally derived interactions in a small world. *Proc. Natl. Acad. Sci.* **100:** 4372–4376.

Goodyer, C.G., Tremblay, J.J., Paradis, F.W., Marcil, A., Lanctot, C., Gauthier, Y., and Drouin, J. 2003. Pitx in vivo promoter activity and mechanisms of positive autoregulation. *Neuroendocrinology* **78:** 129–137.

Gray, P.A., Fu, H., Luo, P., Zhao, Q., Yu, J., Ferrari, A., Tenzen, T., Yuk, D.I., Tsung, E.F., Cai, Z., et al. 2004. Mouse brain organization revealed through direct genome-scale TF expression analysis. *Science* **306:** 2255–2257.

Guelzim, N., Bottani, S., Bourgine, P., and Kepes, F. 2002. Topological and causal structure of the yeast transcriptional regulatory network. *Nat. Genet.* **31:** 60–63.

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.

Han, J.D., Bertin, N., Hao, T., Goldberg, D.S., Berriz, G.F., Zhang, L.V., Dupuy, D., Walhout, A.J., Cusick, M.E., Roth, F.P., et al. 2004. Evidence for dynamically organized modularity in the yeast protein–protein interaction network. *Nature* **430:** 88–93.

Hobert, O. 2005. Specification of the nervous system. In *WormBook* (eds. The *C. elegans* Research Community, WormBook). doi:10.1895/wormbook.1.12.1, http://www.wormbook.org.

Jin, Y., Hoskins, R., and Horvitz, H.R. 1994. Control of type-D GABAergic neuron differentiation by *C. elegans* UNC-30 homeodomain protein. *Nature* **372:** 780–783.

Johnston Jr., R.J., Copeland, J.W., Fasnacht, M., Etchberger, J.F., Liu, J., Honig, B., and Hobert, O. 2006. An unusual Zn-finger/FH2 domain protein controls a left/right asymmetric neuronal fate decision in *C. elegans*. *Development* **133:** 3317–3328.

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.

Kammermeier, L. and Reichert, H. 2001. Common developmental genetic mechanisms for patterning invertbrate and vertebrate brains. *Brain Res. Bull.* **55:** 675–682.

Kong, Y.M., Macdonald, R.J., Wen, X., Yang, P., Barbera, V.M., and Swift, G.H. 2006. A comprehensive survey of DNA-binding transcription factor gene expression in human fetal and adult organs. *Brain Res. Gene Expr. Patterns* **6:** 678–686.

Kummerfeld, S.K. and Teichmann, S.A. 2006. DBD: A transcription factor prediction database. *Nucleic Acids Res.* **34:** D74–D81.

Lanjuin, A. and Sengupta, P. 2004. Specification of chemosensory neuron subtype identities in *Caenorhabditis elegans*. *Curr. Opin. Neurobiol.* **14:** 22–30.

Lanjuin, A., VanHoven, M.K., Bargmann, C.I., Thompson, J.K., and Sengupta, P. 2003. Otx/otd homeobox genes specify distinct sensory neuron identities in *C. elegans*. *Dev. Cell* **5:** 621–633.

Levine, M. and Tjian, R. 2003. Transcription regulation and animal diversity. *Nature* **424:** 147–151.

Luke, G.N., Castro, L.F., McLay, K., Bird, C., Coulson, A., and Holland, P.W. 2003. Dispersal of NK homeobox gene clusters in amphioxus and humans. *Proc. Natl. Acad. Sci.* **100:** 5292–5295.

Melkman, T. and Sengupta, P. 2005. Regulation of chemosensory and GABAergic motor neuron development by the *C. elegans* Aristaless/Arx homolog *alr-1*. *Development* **132:** 1935–1949.

Michaloski, J.S., Galante, P.A., and Malnic, B. 2006. Identification of potential regulatory motifs in odorant receptor genes by analysis of promoter sequences. *Genome Res.* **16:** 1091–1098.

Miller, D.M., Shen, M.M., Shamu, C.E., Burglin, T.R., Ruvkun, G., Dubois, M.L., Ghee, M., and Wilson, L. 1992. *C. elegans unc-4* gene

encodes a homeodomain protein that determines the pattern of synaptic input to specific motor neurons. *Nature* **355:** 841–845.

Miyabayashi, T., Palfeyman, M.T., Sluder, A.E., Slack, F., and Sengupta, P. 1999. Expression and function of members of a divergent nuclear receptor family in *Caenorhabditis elegans*. *Dev. Biol.* **215:** 314–331.

Nelson, C.E., Hersh, B.M., and Carroll, S.B. 2004. The regulatory content of intergenic DNA shapes genome architecture. *Genome Biol.* **5:** R25.

Odom, D.T., Dowell, R.D., Jacobsen, E.S., Nekludova, L., Rolfe, P.A., Danford, T.W., Gifford, D.K., Fraenkel, E., Bell, G.I., and Young, R.A. 2006. Core transcriptional regulatory circuitry in human hepatocytes. *Mol. Syst. Biol.* **2:** 0017.

Page, B.D., Guedes, S., Waring, D., and Priess, J.R. 2001. The *C. elegans* E2F- and DP-related proteins are required for embryonic asymmetry and negatively regulate Ras/MAPK signaling. *Mol. Cell* **7:** 451–460.

Patterson, G.I., Koweek, A., Wong, A., Liu, Y., and Ruvkun, G. 1997. The DAF-3 Smad protein antagonizes TGF-β-related receptor signaling in the *Caenorhabditis elegans dauer* pathway. *Genes & Dev.* **11:** 2679–2690.

Pujol, N., Torregrossa, P., Ewbank, J.J., and Brunet, J.F. 2000. The homeodomain protein CePHOX2/CEH-17 controls antero-posterior axonal growth in *C. elegans*. *Development* **127:** 3361–3371.

Ravasz, E., Somera, A.L., Mongru, D.A., Oltvai, Z.N., and Barabasi, A.L. 2002. Hierarchical organization of modularity in metabolic networks. *Science* **297:** 1551–1555.

Reece-Hoyes, J.S., Deplancke, B., Shingles, J., Grove, C.A., Hope, I.A., and Walhout, A.J.M. 2005. A compendium of *C. elegans* regulatory transcription factors: A resource for mapping transcription regulatory networks. *Genome Biol.* **6:** R110.

Resendis-Antonio, O., Freyre-Gonzalez, J.A., Menchaca-Mendez, R., Gutierrez-Rios, R.M., Martinez-Antonio, A., Avila-Sanchez, C., and Collado-Vides, J. 2005. Modular analysis of the transcriptional regulatory network of *E. coli*. *Trends Genet.* **21:** 16–20.

Rothman, A., Feinstein, P., Hirota, J., and Mombaerts, P. 2005. The promoter of the mouse odorant receptor gene M71. *Mol. Cell. Neurosci.* **28:** 535–546.

Sarafi-Reinach, T.R., Melkman, T., Hobert, O., and Sengupta, P. 2001. The *lin-11* LIM homeobox gene specifies olfactory and chemosensory neuron fates in *C. elegans*. *Development* **128:** 3269–3281.

Satterlee, J.S., Sasakura, H., Kuhara, A., Berkeley, M., Mori, I., and Sengupta, P. 2001. Specification of thermosensory neuron fate in *C. elegans* requires *ttx-1*, a homolog of otd/Otx. *Neuron* **31:** 943–956.

Segal, E., Shapira, M., Regev, A., Pe'er, D., Botstein, D., Koller, D., and Friedman, N. 2003. Module networks: Identifying regulatory modules and their condition-specific regulators from gene expression data. *Nat. Genet.* **34:** 166–176.

Sonnichsen, 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.

Su, A.I., Wiltshire, T., Batalov, S., Lapp, H., Ching, K.A., Block, D., Zhang, J., Soden, R., Hayakawa, M., Kreiman, G., et al. 2004. A gene atlas of the mouse and human protein-encoding transcriptomes. *Proc. Natl. Acad. Sci.* **101:** 6062–6067.

Swoboda, P., Adler, H.T., and Thomas, J.H. 2000. The RFX-type transcription factor DAF-19 regulates sensory neuron cilium formation in *C. elegans*. *Mol. Cell* **5:** 411–421.

Tavazoie, S., Hughes, J.D., Campbell, M.J., Cho, R.J., and Church, G.M. 1999. Systematic determination of genetic network architecture. *Nat. Genet.* **22:** 281–285.

Thieffry, D., Huerta, A.M., Perez-Rueda, E., and Collado-Vides, J. 1998. From specific gene regulation to genomic networks: A global analysis of transcriptional regulation in *Escherichia coli*. *Bioessays* **20:** 433–440.

Tucker, M., Sieber, M., Morphew, M., and Han, M. 2005. The *Caenorhabditis elegans* aristaless orthologue, *alr-1*, is required for maintaining the functional and structural integrity of the amphid sensory organs. *Mol. Biol. Cell* **16:** 4695–4704.

Vazquez, A., Flammini, A., Maritan, A., and Vespignani, A. 2003. Modeling of protein interaction networks. *ComplexUs* **1:** 38.

Walhout, A.J.M. 2006. Unraveling transcription regulatory networks by protein–DNA and protein–protein interaction mapping. *Genome Res.* **16:** 1445–1454.

Walhout, A.J.M., Sordella, R., Lu, X., Hartley, J.L., Temple, G.F., Brasch, M.A., Thierry-Mieg, N., and Vidal, M. 2000a. Protein interaction mapping in *C. elegans* using proteins involved in vulval development. *Science* **287:** 116–122.

Walhout, A.J., Temple, G.F., Brasch, M.A., Hartley, J.L., Lorson, M.A., van den Heuvel, S., and Vidal, M. 2000b. GATEWAY recombinational cloning: Application to the cloning of large numbers of open reading frames or ORFeomes. *Methods Enzymol.* **328:** 575–592.

Woolfe, A., Goodson, M., Goode, D.K., Snell, P., McEwen, G.K., Vavouri, T., Smith, S.F., North, P., Callaway, H., Kelly, K., et al. 2005. Highly conserved non-coding sequences are associated with vertebrate development. *PLoS Biol.* **3:** e7.

Yook, S.H., Oltvai, Z.N., and Barabasi, A.L. 2004. Functional and topological characterization of protein interaction networks. *Proteomics* **4:** 928–942.

Yu, H., Greenbaum, D., Xin Lu, H., Zhu, X., and Gerstein, M. 2004. Genomic analysis of essentiality within protein networks. *Trends Genet.* **20:** 227–231.