Vermeirssen et al.

Supplemental Materials

Methods

A scoring system for high-throughput Y1H data

We developed a standardized scoring system for high-throughput Y1H assays, in which PDIs obtained a score between 1 and 10. Interactions with a higher score are considered of higher confidence. Several factors contribute to the quality of a PDI, including the promoter (DNA bait), the interactor (protein prey) and the PDI itself (Table S3). The promoter affects the Y1H analysis when it exhibits self-activation because it is more difficult to distinguish true from false Y1H positives (Deplancke et al. 2004). A promoter is considered self-active when it confers growth on media containing ≥ 80 mM 3AT (*HIS3*) and/or colored dark blue in the b-galactosidase assay (*lacZ*)(criteria 1 and 2). Two criteria evaluate the influence of an interactor on the quality of a PDI. First, PDIs with interactors that possess a known DNA binding domain are considered more reliable (criterion 3). Second, we found that several interactors (CEY-2, EGL-44, TAB-1, F26H9.2, H02I12.5, NCX-8, PRX-5 and RHR-1) were more likely to give a positive Y1H readout for the *HIS3* reporter only (81%), compared to other interactors (8%, t-test, P <0.001). PDIs involving these "sticky on *HIS3*" interactors may constitute false positives when *lacZ* reporter activation cannot be analyzed and are, therefore, penalized in the scoring system. Similarly, although this was not observed in this dataset, interactors could be identified that have a higher tendency to generate

lacZ reporter expression than normal and these could contribute to false positives when the *HIS3* reporter activation cannot be analyzed (criterion 4). Several criteria assessed the quality of the PDI itself. PDIs are more reliable if they were found multiple times in screens compared to only once (77% *versus* 49%, Chi-square, P <0.001, criterion 5). PDIs that were found only once, but that involve a predicted TF can be grouped in two categories: those found from the AD-wrmcDNA library and those found from the AD-TF mini-library. Obviously, the first category is of higher confidence than the second, as TFs only make up about 5% of *C. elegans* protein-coding genes (criterion 6). Finally, the more independent experiments in which a PDI is detected, the higher is its likelihood of being a true PDI (criteria 7-10). A total score between 1 and 10 was obtained for each PDI by dividing the weighted sum for all criteria by the sum of all the weights for the criteria that applied and multiplying it by a factor of 10. PDIs were grouped in 10 different score categories (Fig. S2). Starting off with a weight of 1 for all criteria, weights and score cut-off were optimized to obtain the best extraction of a high confidence dataset. The weights were varied from 1 to 3, with increments of 1. We considered a set of known high- and low-quality PDIs. As high-quality interactions, we defined a set of 120 PDIs that were retrieved in screens/mating and that were confirmed in the Y1H matrix experiment. As low quality interactions, we defined a set of 79 PDIs: 1) PDIs for which the promoters' *lacZ* reporter gene was self-active or absent and the interactor was sticky on *HIS3*, 2) PDIs that were not confirmed in the Y1H matrix experiment and found only once from AD-TF library, 3) PDIs that were not confirmed and found only

once from AD-wrmcDNA library or mating with novel interactors or interactor TFs for which was known that their corresponding ORFeome clones were functional in the Y1H matrix experiment. An interactor clone was regarded as functional, if it was able to confirm PDIs found from the AD-wrmcDNA library or was found in a new PDI and its identity was confirmed by sequencing. The low and high quality interactions were used as a guidance to place the score cut-off. A cut-off of 5 created a dataset that retained all high quality interactions (100%) and overall most PDIs (75%), while removing most low quality interactions (95%). All PDIs and their respective scores are depicted in Table S2. All previously reported regulatory interactions and regulogs in Fig. 2 had a score higher than 5 (Table S2).

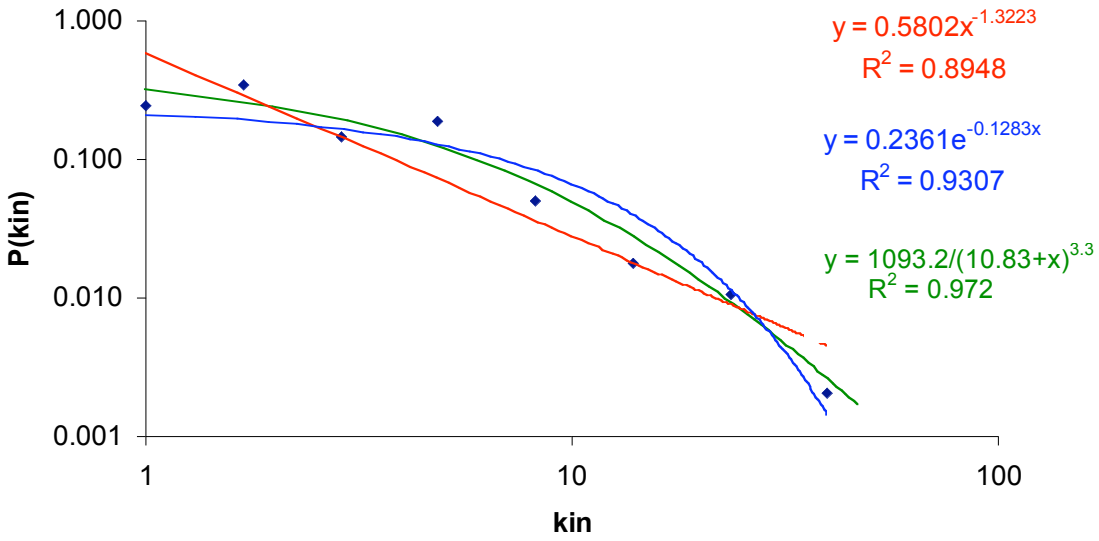
Network analysis

Y1H PDIs were visualized into a network using Cytoscape v2.3 (Shannon et al. 2003). Different promoter variants for one gene were combined into a single node. As a measure of connectivity, the number of pairs of nodes with a directed/undirected path between them was calculated using Pajek (de Nooy et al. 2005). The in-degree k_{in} is the number of incoming links per node, *i.e.* the number of TFs that bind a promoter. The out-degree k_{out} is the number of outgoing links per node, *i.e.* the number of promoters that a TF binds to. $<k_{in}>$ and $<k_{out}>$, are the averages of the values over all the nodes in the network. Logarithmic binning was used to fit power law curves to $P(k_{in})$, the incoming connectivity, and $P(k_{out})$, the outgoing connectivity. The clustering coefficient of a

node, C, is defined as the ratio of the number of existing links between a node's interaction partners and the maximum possible number of links between these neighbors, which can be calculated for every node that has more than one link. The clustering coefficient of a network <C> is the average of all the individual clustering coefficients. For the calculation of <C>, the directionality of the links was not taken into account. The average clustering coefficient of a random network of similar size and degree is given by the average degree of the network <k> divided by the total number of nodes (Albert and Barabasi 2002). The average clustering coefficient was also calculated for 100 randomized networks with the same single node characteristics, generated by shuffling the links 1000 times. To uncover the underlying modularity of the network, a topological overlap matrix was created for promoters and interactors, respectively, by calculating the (directed) topological overlap coefficient (TOC) for every node that had more than 1 link. The TOC or mutual clustering coefficient is a relative measure of the number of neighboring nodes that are shared between two nodes. Both the meet/min and the geometric formula were used (Goldberg and Roth 2003). Subsequently, an average-linkage hierarchical clustering algorithm was applied to the topological overlap matrix, which placed the nodes with a similar topological overlap to other nodes, close to each other (Eisen et al. 1998). The modularity was also visualized by plotting the topological overlap network, which linked nodes with a certain TOC value, in an organic layout in Cytoscape v2.3. The organic layout algorithm allows the visualization of the clustered structure of a graph, *i.e.* groups of highly interconnected nodes (Cytoscape manual).

Degree distribution analysis

The in –and outgoing degree distribution were fitted to three different curves, including a power law, a power law with saturation and an exponential curve. The fits for the incoming degree distribution are shown below. The R^2 values are indicated in the Table.



In Degree	Power-Law	Power-Law with Saturation	Exponential
R2	0.89	0.97	0.93
Max. residual	0.127	0.018	0.030

Out-Degree	Power-Law		Exponential
R2	0.99	Not applicable	0.88
Max. residual	0.0015	Not applicable	0.2534

C32D5.1

AD-C32D5.1 was transformed into 8 digestive tract promoter bait strains and in 8 neuronal promoter bait strains as control. Chromatin immunoprecipitation in yeast was performed using a *Pcog-1* bait strain with an anti Gal4-AD antibody as described (Deplancke et al. 2006a). PCR promoter scanning in fragments of ~200 bp was performed.

Mouse expression profile analysis

Mouse expression data were obtained from SymAtlas (Su et al. 2004). A Fisher test was used to test if there was an association between the presence of a particular DNA binding domain in a TF and the expression of that TF in a specific mouse tissue. The following tissues were examined: heart, endocrine tissue (adrenal gland, pancreas, pituitary, thyroid), fat tissue (adipose tissue, brown fat), epithelial system (epidermis, snout epidermis, tongue epidermis), digestive tract (large intestine, small intestine, salivary gland, stomach), bone and bone marrow, muscle, neuronal tissue (amygdale, cerebellum, cerebral cortex, cortex, dorsal root ganglion, dorsal striatum, frontal cortex, hippocampus, hypothalamus, medial olfactory epithelium, olfactory bulb, pre-optic, retina, spinal cord lower, spinal cord upper, substantia nigra, trigeminal, vomeral nasal organ), reproductive system (ovary, testis, uterus, prostate), respiratory system (trachea, lung), immune system (b220+bcell, cd4+Tcell, cd8+Tcell, thymus, lymph node, spleen), liver, excretory system (bladder, kidney) and other (digits, mammary gland lactating, placenta, umbilical cord). From the 1305 mouse TFs, 774 TFs were represented, and for 582 of these an expression profile could be extracted.

These TFs had a total of 655 different DNA binding domains. Expression information was retrieved for 46.1% of all DNA binding domains, including 46.3% of homeodomains, 63.8% of basic helix-loop-helices, 52.6% of winged helices, 59.2 % bZIPs, 42.0% of C2H2 zinc fingers and 59% of nuclear hormone receptors. Briefly, for each feature on the array the Affymetrix Microarray Suite 5.0 software had generated a present (P) marginal (M) or absent (A), and an intensity value. Features close to the limit of detection often had a P call, but an intensity value smaller than the value of some of the features identified as A. To avoid including those P features in our analysis, a gene was considered to be expressed in a tissue if it had a P call and if the intensity of the feature was higher than the mean of intensities of the A features in that sample plus two times the standard deviation in at least one of the replicas. The calculations for mouse DNA binding domains were calculated using two other cut-offs: 1) calling a gene present if it had a P call in both replicas and 2) calling a gene present if it had a P call and intensity higher than the mean plus two times the standard deviation for A features in both replicas. In both cases, a significant association between homeodomain and neurons was obtained using a Fisher test.

Supplemental Figure legends

Figure S1 Yeast-one hybrid (Y1H) pipeline.

Initially all promoter baits were subjected to AD-wrmcDNA and AD-TF mini-library screens. In addition, some baits, mostly baits for which no positives were

retrieved from the screens or baits that were highly self-active, were also analyzed by mating (Table S2). Finally, all interactions obtained by these three Y1H methods were confirmed in a matrix experiment, which tested all yeast promoter strains against all interactors obtained and interactor TFs encoded by the target genes (Tables S2 and S3).

Figure S2 A standardized Y1H scoring system.

Evaluation of the Y1H scoring system for the weights specified in Table S3. PDIs were divided in 10 score categories: score category 10 contained PDIs with a score equal to 10; score category 9 contained PDIs with a score equal to 9 or higher, but lower than 10; score category 8 contained PDIs with a score equal to 8 or higher, but lower than 9; *etcetera*; score category 1 contained PDIs with a score equal to 1 or higher, but lower than 2. (*A*) Bar graph representing the complete PDI dataset. Known low and high-quality PDIs (Methods) are indicated. Total – total number of PDIs per score category. (*B*) A cut-off score of 5 or higher was chosen to obtain a high-quality PDI dataset, since at this score cut-off the percentage of high-quality interactions included and the percentage of low-quality interactions excluded, was the highest.

Figure S3 C32D5.1, a novel putative global regulator. (*A*) C32D5.1 can bind a similar proportion of digestive tract gene promoters as neuronal gene promoters. Y1H experiment using AD-C32D5.1 or AD alone. Left panel – permissive media, middle panel – selective media, right panel – bGal assay. N – neuronal target

genes, DT – digestive tract target genes. Arrows indicate double positives. (*B*) C32D5.1 is a novel putative DNA binding protein. ChIP assay in a *Pcog-1* yeast strain transformed with AD-C32D5.1 using an anti Gal4AD antibody. The samples were treated with Gal4AD antibody during chromatin immunoprecipitation to select for specific AD-TF/DNA complexes, whereas the inputs contained sheared genomic yeast DNA (not precipitated). AD input/sample refers to yeast transformed with an empty pAD-DEST plasmid, i.e. that does not contain any TF. Samples and inputs (not precipitated) were analyzed by PCR with primers scanning the promoter in fragments of ~200 bp. Fragment 5, 6, 7 and 8 (bold) were clearly detected in the sample and not in the AD alone. The picture illustrates the binding to fragment 5 (red). Black triangles indicate two-fold titrations of input and sample. (*C*) Alignment of Y55F3BR.5 and C32D5.1. BestFit was used to create the alignment. * identical amino acids, : conservation of strong groups, . conservation of weak groups.

Figure S4 The core neuronal PDI network.

Dark blue diamonds – promoters; orange circles – interactors; green triangles – interactors whose promoters were also analyzed in this network; yellow circles and triangles – interactor hubs; light blue diamonds and triangles – promoter hubs.

Figure S5 TF modules as determined by the meet/min formula.

Supplemental tables

Table S1: Network nodes information

Table S2: Bait information

Table S3: Interaction scoring matrix

Table S4: Y1H scoring system

Table S5: Network properties

Table S6: Module info

Table S7: Primer sequences

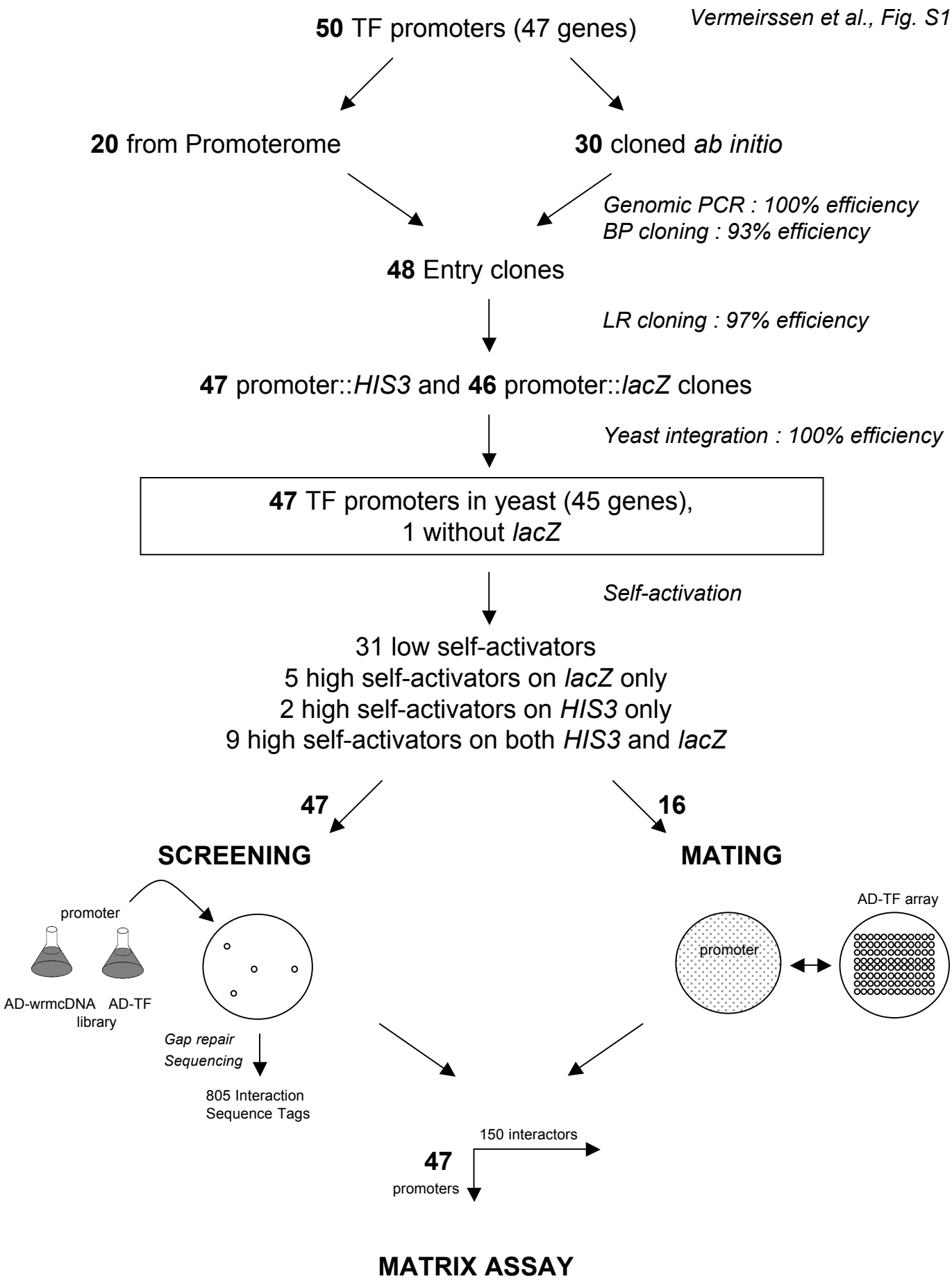
Table S8: Nominal and Bonferroni corrected P-values for GO analysis

REFERENCES

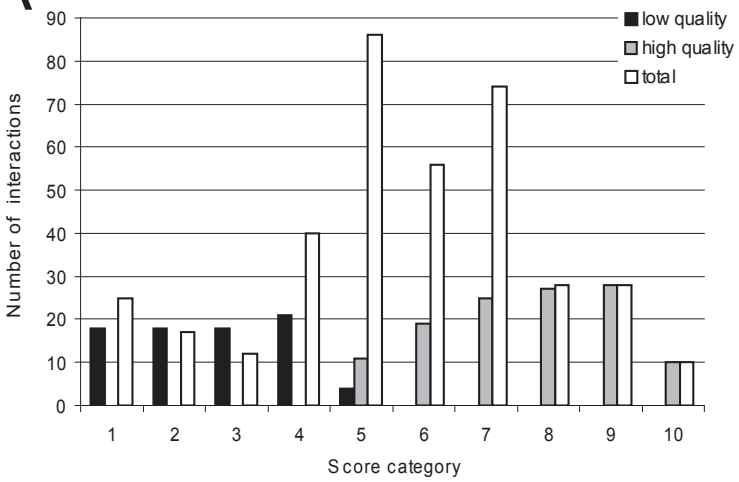
- Albert, R. and A.L. Barabasi. 2002. Statistical mechanics of complex networks. *Rev. Mod. Phys.* **74**: 47-97.
- Deplancke, B., D. Dupuy, M. Vidal, and A.J.M. Walhout. 2004. A Gateway-compatible yeast one-hybrid system. *Genome Res* **14**: 2093-2101.
- Deplancke, B., A. Mukhopadhyay, W. Ao, A.M. Elewa, C.A. Grove, N.J. Martinez, R. Sequerra, L. Doucette-Stam, J.S. Reece-Hoyes, I.A. Hope, H.A. Tissenbaum, S.E. Mango, and A.J.M. Walhout. 2006a. A gene-centered *C. elegans* protein-DNA interaction network. *Cell* **125**: 1193-1205.
- Eisen, M.B., P.T. Spellman, P.O. Brown, and D. Botstein. 1998. Cluster analysis and display of genome-wide expression patterns. *Proc. Natl. Acad. Sci. USA* **95**: 14863-14868.

- Goldberg, D.S. and F.P. Roth. 2003. Assessing experimentally derived interactions in a small world. *Proc Natl Acad Sci U S A* **100**: 4372-4376.
- Shannon, P., A. Markiel, O. Ozier, N.S. Baliga, J.T. Wang, D. Ramage, N. Amin, B. Schwikowski, and T. Ideker. 2003. Cytoscape: a software environment for integrated models of biomolecular interaction networks. *Genome Res* **13**: 2498-2504.
- Su, A.I., T. Wiltshire, S. Batalov, H. Lapp, K.A. Ching, D. Block, J. Zhang, R. Soden, M. Hayakawa, G. Kreiman, M.P. Cooke, J.R. Walker, and J.B. Hogenesch. 2004. A gene atlas of the mouse and human protein-encoding transcriptomes. *Proc Natl Acad Sci U S A* **101**: 6062-6067.

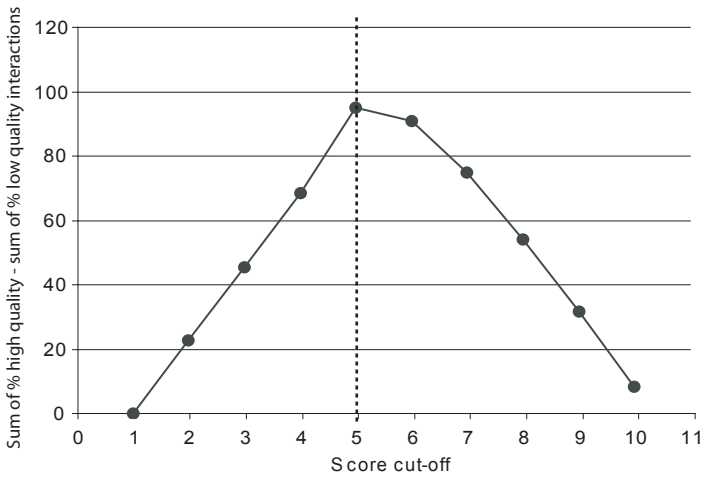
Vermeirssen et al., Fig. S1



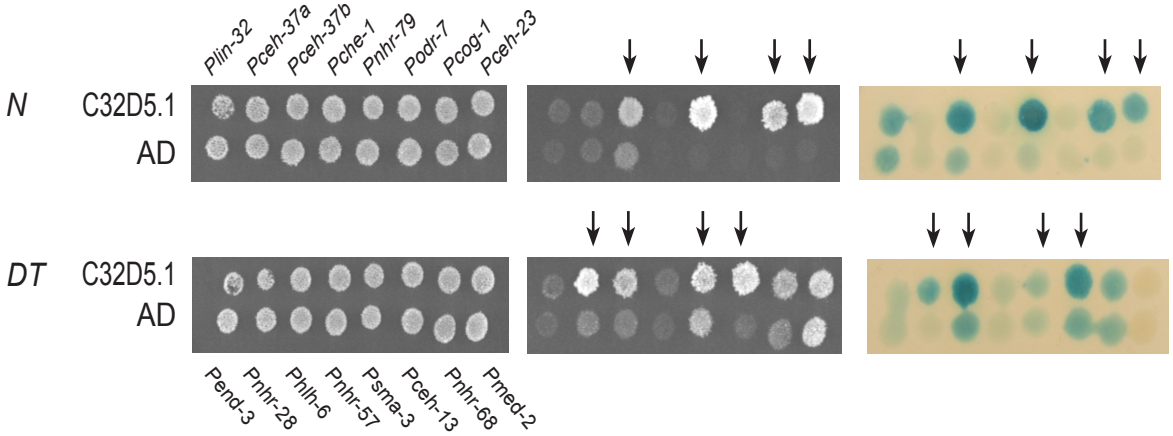
A


low quality
high quality
total
Number of interactions
Score category
0
10
20
30
40
50
60
70
80
90
1
2
3
4
5
6
7
8
9
10

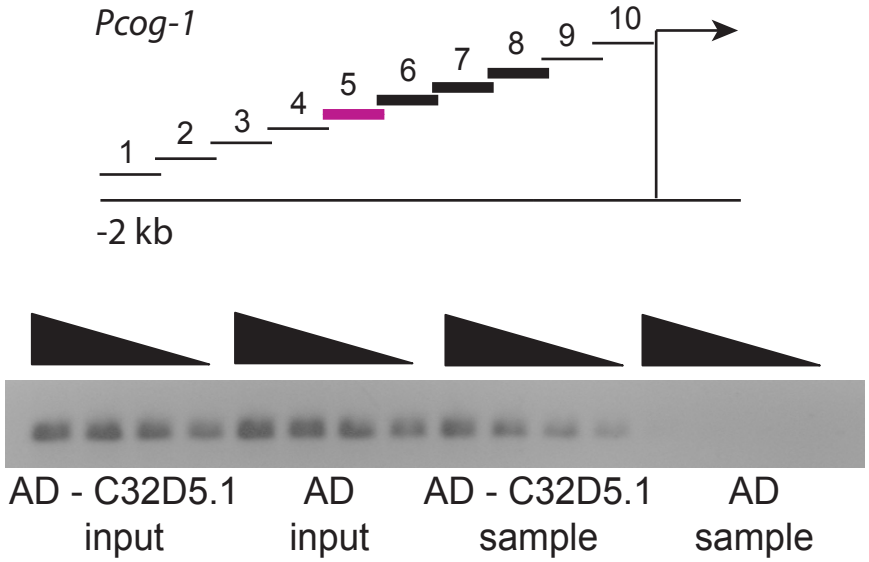
B


Sum of % high quality - sum of % low quality interactions
Score cut-off
0
20
40
60
80
100
120
0
1
2
3
4
5
6
7
8
9
10
11

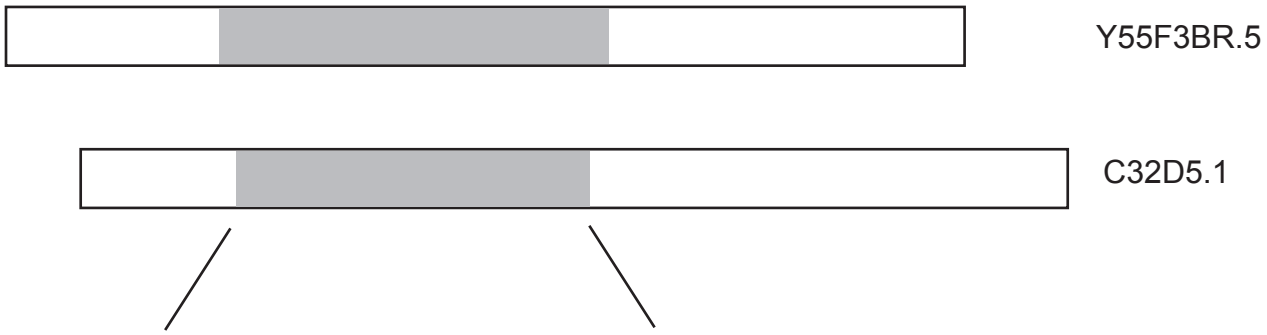
A



B



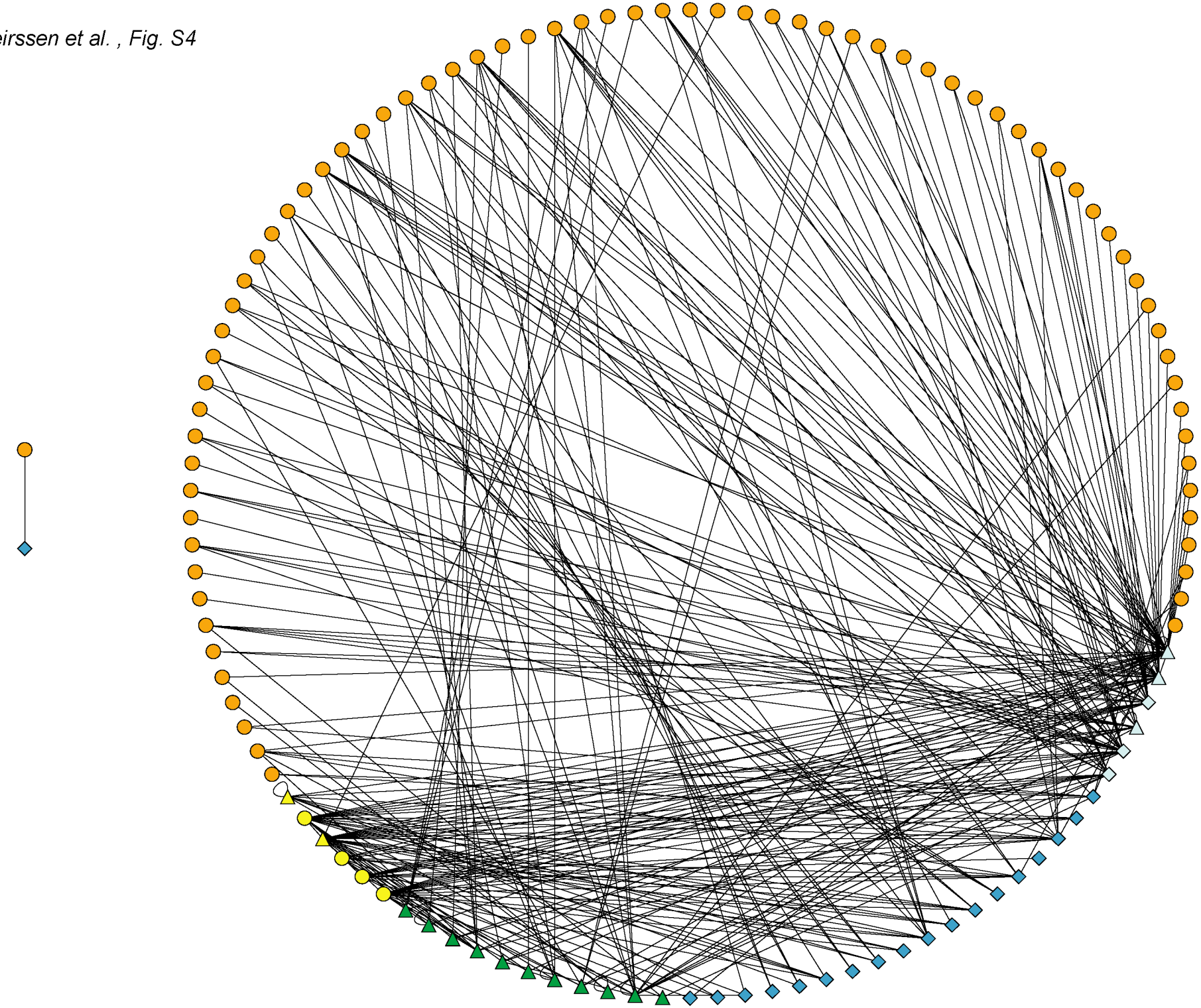
C




```
Y55F3BR.5_60-155_      FKCYEERDIEFFIHQVSDNQIIWDRRNSLNFDGNLVHATFDKIEICC-QFLARKNGENAS
C32D5.1__41-130_       ----------FFISLFSGREELYSP---CKKNSDFRQKAFYIIEKKCGHFLAMRKGRNAE
                              ***  .*..: ::.     : :.:: : :*  **  * :*** ::*.**.

Y55F3BR.5_60-155_      RLWTELASEYK------REKARIEQLPSGIYSEEFEIRFPFMQQMWFLEIEAS
C32D5.1__41-130_       KLWLYLFNDFEKMIRNGGDKNDEKSKSIVPFYDCLQFLIPYLEKTDSLSKLTL
                       :**  * .:::       :*   :. .   : : ::: :*::::   *.  :
```

Vermeirssen et al. , Fig. S4



Vermeirssen et al. , Fig. S5

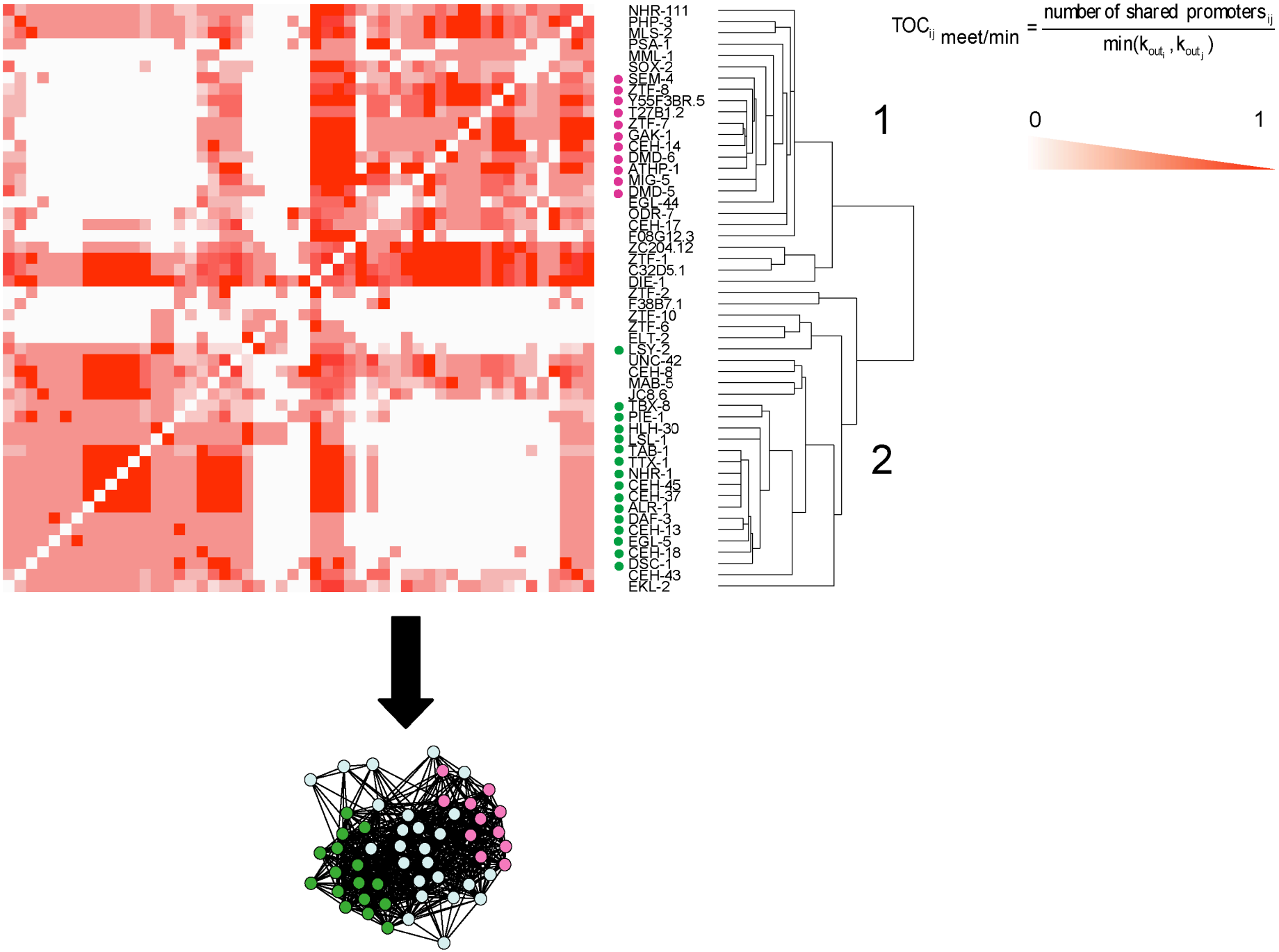


Table S2

Sequence name	Public name	Prom::HIS3 available	Prom::lacZ available	HIS3 self-activation	lacZ self-activation	mating	interactions found?
PR08B4.2	*Palr-1*	Yes	Yes			No	No
PC33A11.4	*PC33A11.4*	Yes	Yes	high	high	Yes	Yes
PW03A3.1	*Pceh-10*	Yes	Yes		high	No	Yes
PF46C8.5	*Pceh-14*	Yes	Yes			Yes	Yes
PZK652.5	*Pceh-23*	Yes	Yes			No	Yes
PC33D12.7	*Pceh-30*	Yes	Yes			No	No
PW05E10.3	*Pceh-32*	Yes	Yes			No	No
PC37E2.4	*Pceh-36*	Yes	Yes	high	high	Yes	Yes
PC37E2.5	*Pceh-37a*	Yes	Yes			No	Yes
PC37E2.6	*Pceh-37b*	Yes	Yes			No	Yes
PC28A5.4	*Pceh-43*	Yes	Yes			No	Yes
PC55B7.12	*Pche-1*	Yes	Yes			Yes	Yes
PR03C1.3	*Pcog-1*	Yes	Yes			No	Yes
PB0412.1	*Pdac-1*	Yes	Yes			No	Yes
PF11A1.3	*Pdaf-12a*	No	No		N/A	N/A	N/A
PF11A1.3	*Pdaf-12b*	Yes	No		N/A	No	Yes
PR13H8.1	*Pdaf-16a*	Yes	Yes	high		Yes	Yes
PR13H8.1	*Pdaf-16b*	Yes	Yes			No	Yes
PF33H1.1	*Pdaf-19*	Yes	Yes		high	No	Yes
PF25E2.5	*Pdaf-3*	Yes	Yes			Yes	Yes
PC18D1.1	*Pdie-1*	Yes	Yes			No	Yes
PR53.3	*Pegl-43*	Yes	Yes			No	Yes
PT14G12.4	*Pfkh-2*	Yes	Yes			Yes	No
PK01B6.1	*Pfozi-1*	N	N		N/A	N/A	N/A
PM05B5.5	*Phlh-2*	Yes	Yes		high	No	Yes
PZC64.4	*Plim-4*	Yes	Yes			Yes	Yes
PK03E6.1	*Plim-6*	Yes	Yes			No	Yes
PZC247.3	*Plin-11*	Yes	Yes			Yes	Yes
PT14F9.5	*Plin-32*	Yes	Yes			No	Yes
PF42A9.2	*Plin-49*	Yes	Yes			No	Yes
PW10D5.1	*Pmef-2*	Yes	Yes	high		No	Yes
PF07C3.10	*Pnhr-36*	No	No		N/A	N/A	N/A
PK01H12.3	*Pnhr-38*	Yes	Yes			No	Yes
PY104H12A.1	*Pnhr-41*	Yes	Yes	high	high	No	Yes
PC48D5.1	*Pnhr-6*	Yes	Yes	high	high	No	Yes
PT26H2.9	*Pnhr-79*	Yes	Yes			Yes	Yes

PF48G7.3	*Pnhr-83*	Yes	Yes	high	high	Yes	Yes
PT18D3.2	*Podr-7*	Yes	Yes			No	Yes
PT19E7.2	*Pskn-1*	Yes	Yes		high	No	No
PT22H9.4	*PT22H9.4*	Yes	Yes	high	high	No	Yes
PF21H11.3	*Ptbx-2*	Yes	Yes	high	high	Yes	Yes
PY113G7A.6	*Pttx-1*	Yes	Yes			No	Yes
PC40H5.5	*Pttx-3*	Yes	Yes			Yes	No
PC47G2.2	*Punc-130*	Yes	Yes			No	Yes
PY16B4A.1	*Punc-3*	Yes	Yes	high	high	Yes	Yes
PB0564.10	*Punc-30*	Yes	Yes		high	Yes	Yes
PW02D3.9	*Punc-37*	Yes	Yes			No	Yes
PF58E6.10	*Punc-42*	Yes	Yes	high	high	Yes	Yes
PT28F12.2	*Punc-62*	Yes	Yes			No	Yes
PC30A5.7	*Punc-86*	Yes	Yes			No	Yes

Table S4. Y1H scoring system based on 10 different criteria. A total score between 1 and 10 was obtained for each PDI by dividing the weighted sum for all criteria by the sum of all the weights for the criteria that applied and multiplying it by a factor of 10.

Scoring of		Scoring criterium	Score 0	Score 1	Weight	Weighted score
PROMOTER (Y1H bait)	1	*lacZ* self-active or absent?	yes	no	1	$score_1$ x $weight_1$
	2	*HIS3* self-active?	yes	no	1	$score_2$ x $weight_2$
INTERACTOR (Y1H prey)	3	Predicted TF (wTF2.1)?	no	yes	1	$score_3$ x $weight_3$
	4	Sticky on *HIS3/lacZ*?	yes	no	3	$score_4$ x $weight_4$
INTERACTION in screens	5	Found multiple times?	no	yes	1	$score_5$ x $weight_5$
	6	Found only once from AD-TF library?	yes	no	1	$score_6$ x $weight_6$
	7	Found by both libraries?	no	yes	1	$score_7$ x $weight_7$
INTERACTION in mating	8	Detected by mating?[a]	no	yes	1[a]	$score_8$ x $weight_8$*
INTERACTION in matrix assay	9	*HIS3* positive?[b]	no	yes	2[b]	$score_9$ x $weight_9$*
	10	*lacZ* positive?[c]	no	yes	2[c]	$score_{10}$ x $weight_{10}$*
Total score						$\frac{Sum\ of\ weighted\ scores}{Sum\ of\ weights} \times 10$

[a] only applicable if promoter was tested in mating

[b] only applicable if PDI was tested in matrix assay

[c] only applicable if PDI was tested in matrix assay and *lacZ* reporter was present and not self-active

Table S5. Topological properties of the core neuronal PDI network. Different promoter variants for one gene were combined into a single node. Both *Pceh-37a* and *Pceh-37b* were bound by DIE-1, resulting in only 1 directed link between the nodes DIE-1 and CEH-37. Hence, 282 overall PDIs, but only 281 in the network.

Network parameter	**Neuronal core PDI network**
# target genes	38
# interactors	94
# nodes	116
# links	281
# pairs with undirected path	12884 (96.6%)
# pairs with directed path	434 (3.3%)
$<k_{in}>$	7.40
median k_{in}	4
$<k_{out}>$	2.99
median k_{out}	2
<C>	0.132 ± 0.026
<C> random network	0.04
<C> randomized network	0.115 ± 0.002

Table 6 Module annotation

	Degree	Expression	Gene Ontology	DNA binding domain	RNAi or mutant phenotype	Functional indicator based on GO and phenotype
M1 Interactors						
ZTF-8	5	Neurons and elsewhere	gametogenesis; locomotory behavior; physiological process	ZF - C2H2	embryonic development abnormal; postembryonic development abnormal; Stp; Sck; locomotion abnormal	BR
DMD-5	4	Only in neurons	hermaphrodite genitalia development; locomotory behavior; morphogenesis of an epithelium; oviposition; sex differentiation	ZF - DM	Rup, Pvl, Egl; locomotion abnormal	BR
MIG-5	4	Neurons and elsewhere	cell migration; embryonic cleavage; embryonic development; morphogenesis of an epithelium; oviposition; spindle organization and biogenesis; development; frizzled signaling pathway; intracellular signaling cascade	WH	Emb, Let, Egl, Rup, Mig	BR
DMD-6	3	NA	sex differentiation	ZF - DM	NA	_
Y55F3BR.5	3	NA	NA	MADF	NA	_
SEM-4	3	Neurons and elsewhere	oviposition	ZF - C2H2	Egl	BR
T27B1.2	3	NA	embryonic development; growth; larval development; locomotory behavior; physiological process; positive regulation of body size; post-embryonic body morphogenesis	ZF - C2H2	Let, Lva, Bmd, Dpy; locomotion abnormal	B
ZTF-7	2	Not in neurons	NA	ZF - C2H2	NA	_
ATHP-1	2	NA	reproduction	AT hook	maternal sterile; embryonic development abnormal; postembryonic development abnormal	R
GAK-1	2	Not in neurons	reproduction	AT hook	maternal sterile; embryonic development abnormal; postembryonic development abnormal; reduced brood size	R
CEH-14	2	Neurons and elsewhere	thermosensory behavior	HD - LIM	NA	B

	Degree	Expression	Gene Ontology	DNA binding domain	RNAi or mutant phenotype	Functional indicator based on GO and phenotype
M1 Promoters						
daf-3	27	Neurons and elsewhere	NA	*	Dauer	B
cog-1	25	Neurons and elsewhere	reproduction	*	Ste, Egl, vulval development	R
nhr-79	22	Neurons and elsewhere	NA	*	NA	_

	Degree	Expression	Gene Ontology	DNA binding domain	RNAi or mutant phenotype	Functional indicator based on GO and phenotype
Connector Interactors						
C32D5.1	20	NA	NA	Novel	NA	_
ZTF-1	18	Neurons and elsewhere	NA	ZF - C2H2	embryonic development abnormal, postembryonic development abnormal	_
DIE-1	18	Neurons and elsewhere	embryonic development; larval development; morphogenesis of embryonic epithelium; negative regulation of vulval development; physiological process; positive regulation of growth rate; post-embryonic body morphogenesis; actin filament organization	ZF - C2H2	Emb, Muv, Bmd, Gro	_
ZC204.12	12	Neurons and elsewhere	NA	Novel	NA	_
JC8.6	9	Neurons and elsewhere	embryonic development; hermaphrodite genitalia development; positive regulation of growth rate	Novel	Pvl, Gro, Let	_
ODR-7	9	Only in neurons	positive chemotaxis	ZF - NHR	NA	B
CEH-17	8	Only in neurons	axon guidance; regulation of axon extension	HD - PRD	NA	_
CEH-8	8	NA	pattern specification	HD - PRD	NA	_

MAB-5	7	Neurons and elsewhere	localization; positive regulation of epithelial cell proliferation; regulation of cell fate specification; regulation of cell migration; tail tip morphogenesis; anterior/posterior pattern formation	HD - HOX	embryonic development abnormal, postembryonic development abnormal	_
UNC-42	7	Neurons and elsewhere	locomotory behavior	HD - PRD	locomotion abnormal	B
PHP-3	6	NA	NA	HD - HOX	embryonic development abnormal, postembryonic development abnormal	_
EKL-2	6	Neurons and elsewhere	embryonic development; locomotory behavior	ZF - C2HC	Let, locomotion abnormal	B
MLS-2	5	Neurons and elsewhere	NA	HD - NK	NA	_
NHR-111	4	Neurons and elsewhere	NA	ZF - NHR	NA	_
CEH-43	3	Neurons and elsewhere	NA	HD - NK	embryonic development abnormal, postembryonic development abnormal, Let, Lva, locomotion abnormal, maternal sterile	BR
EGL-44	3	Neurons and elsewhere	NA	TEA/ATTS	NA	_

	Degree	Expression	Gene Ontology	DNA binding domain	RNAi or mutant phenotype	Functional indicator based on GO and phenotype
M2 Interactors						
TBX-8	4	Not in neurons	embryonic development; growth; larval development; physiological process; post-embryonic body morphogenesis	T-box	Let, Lvl, Bmd	_
LSY-2	6	Neurons and elsewhere	gametogenesis; hermaphrodite genitalia development; locomotory behavior; positive regulation of growth rate	ZF - C2H2	Stp, Loc, Pvl (neuronal symmetry)	BR
PIE-1	4	Not in neurons	cell fate determination; embryonic development; formation of primary germ layer; embryonic pattern specification	ZF - CCCH	embryonic development abnormal, postembryonic development abnormal, Let	_
TAB-1	3	Neurons only	NA	HD - NK	NA	_

TTX-1	2	Neurons and elsewhere	thermosensory behavior	HD - PRD	NA	B
CEH-37	2	Neurons and elsewhere	cell fate specification	HD - PRD	NA	_
ALR-1	2	Neurons and elsewhere	NA	HD - PRD	NA	_
CEH-45	2	Not in neurons	NA	HD - PRD	NA	_
NHR-1	2	Not in neurons	NA	ZF - NHR	NA	_
DSC-1	2	Neurons and elsewhere	defecation	HD -PRD	constipated (Con); expulsion abnormal; short defecation cycle	_
CEH-13	2	Neurons and elsewhere	cell-cell adhesion; embryonic development; growth; larval development; locomotory behavior; positive regulation of body size; post-embryonic body morphogenesis	HD - HOX	embryonic development abnormal, postembryonic development abnormal, Bmd; Dpy; Lva; small	B
DAF-3	2	Neurons and elsewhere	NA	MH1	Dauer	B
LSL-1	2	NA	embryonic development	ZF - C2H2	Emb	_
EGL-5	2	Neurons and elsewhere	anterior/posterior pattern formation; oviposition	HD - HOX	Egl, Dev	BR
HLH-30	2	Neurons and elsewhere	NA	bHLH	Fat	_
CEH-18	2	Not in neurons	epidermis development; germ cell migration; oocyte maturation; ovulation	HD - POU	gonadal sheath cell differentiation; oogenesis defects; larval lethality	R

	Degree	**Expression**	**Gene Ontology**	**DNA binding domain**	**RNAi or mutant phenotype**	**Functional indicator based on GO and phenotype**
M2 promoters						
unc-30	36	Neurons only	NA	*	Unc	B
ceh-23	16	Neurons and elsewhere	NA	*	embryonic development abnormal, postembryonic development abnormal, AIY differentiation	_
nhr-83	10	Neurons only	NA	*	NA	_
T22H9.4	7	Neurons only	sex differentiation	*	NA	_
nhr-41	6	Neurons and elsewhere	NA	*	Dauer	B
ttx-1	5	Neurons and elsewhere	thermosensory behavior	*	NA	B
daf-19	4	Neurons only	NA	*	Dauer	B

hlh-2	4	Neurons and elsewhere	embryonic development; gametogenesis; hermaphrodite genitalia development; larval development; locomotory behavior; physiological process; post-embryonic body morphogenesis; regulation of cell fate specification	*	embryonic development abnormal, postembryonic development abnormal, Bmd, Stp, Pvl, Let, locomotion abnormal	BR
nhr-6	3	Neurons and elsewhere	NA	*	embryonic development abnormal, postembryonic development abnormal, Ovulation abnormal	R
lin-32	2	Neurons only	localization; neuron development	*	NA	_

* only for TF interactors
NA = not available
B = response to stimulus/behavior
R = reproduction

Table S7. Primer sequences for PCR promoter scanning after yeast chromatin immunoprecipitation.

Pcog-1 scanning starting 1997 bp upstream of ATG	Primers used
1	>FW_CCAGGTTCCTTAAAGGTTTCATTGT >RV_AATTATAGGCTTTTTGAACTAAAAAATTTG
2	>FW_AATTTGGCAGAATATTTTAATCTTTCAATG >RV_GAAAACGTCTCAAAAAAGTAGACA
3	>FW_GTCTACTTTTTTGAGACGTTTTCATCTT >RV_ATAGGGTGAGTAAAATTTTGCCGA
4	>FW_TTGAAAATTCCAGAATTTGAATTTAAATCGG >RV_TTCGGCATAATTTTTGTGGGCGAATA
5	>FW_GCCGAACGGCAATTGGCG >RV_TGGAATCTAAGAAAATAGTAATAA
6	>FW_CTTAGATTCCAAGTTAATTCTTGGT >RV_CAGTCTCACTAATACGTCTCTCTCTCA
7	>FW_GAGAGACGTATTAGTGAGA >RV_GGCATTTCAAGATAATAATAGT
8	>FW_ATGCCAAATTATTCTTCATAAGTGCC >RV_TGAATATCTTAAAAATTCCTAGAATTTCG
9	>FW_AAGATATTCAAAAACAGTTTTACCCA >RV_TTAGAAAAAGTGATAATTATTAAATTTCTAAA
10	>FW_TTTAATAATTATCACTTTTTCTAAAATTTAT >RV_CTGGTTATGGTAGAGGGGAGATTGTT

Table S8

	Target genes		Interactor TFs		Module 1**		Module 2**	
	P-value	Bonferroni corrected P-value	P-value	Bonferroni corrected P-value*	P-value	Bonferroni corrected P-value	P-value	Bonferroni corrected P-value
Development	0.0321	0.2245	**0.0041**	**0.0287**	0.4769	1.0000	0.1717	1.0000
Response to stimulus	**0.0004**	**0.0030**	**0.0002**	**0.0011**	**0.0007**	**0.0050**	0.2942	1.0000
Reproduction	0.1132	0.7922	**0.0057**	**0.0401**	**0.0004**	**0.0029**	0.4862	1.0000
Cellular process	0.4115	1.0000	0.1806	1.0000	**0.0022**	**0.0153**	1.0000	1.0000
Growth	0.6223	1.0000	0.7291	1.0000	1.0000	1.0000	0.7292	1.0000
Physiological process	0.5653	1.0000	0.6805	1.0000	0.0574	0.4019	0.6785	1.0000
Regulation of biological process	1.0000	1.0000	0.1689	1.0000	**0.0019**	**0.0130**	0.4245	1.0000

*We checked seven Biological Process Ontology terms for the complete network. Thus we multiplied the P-values by 7 to correct for multiple hypothesis testing (Bonferroni corrected P-values).
**For the modules, we only looked at the Biological Process Ontology terms that were significant for the whole network (n=3). Hence, we multiplied the nominal P-values by 3 to correct for multiple hypothesis testing (Bonferroni corrected P-values).
Bold indicates significant values.