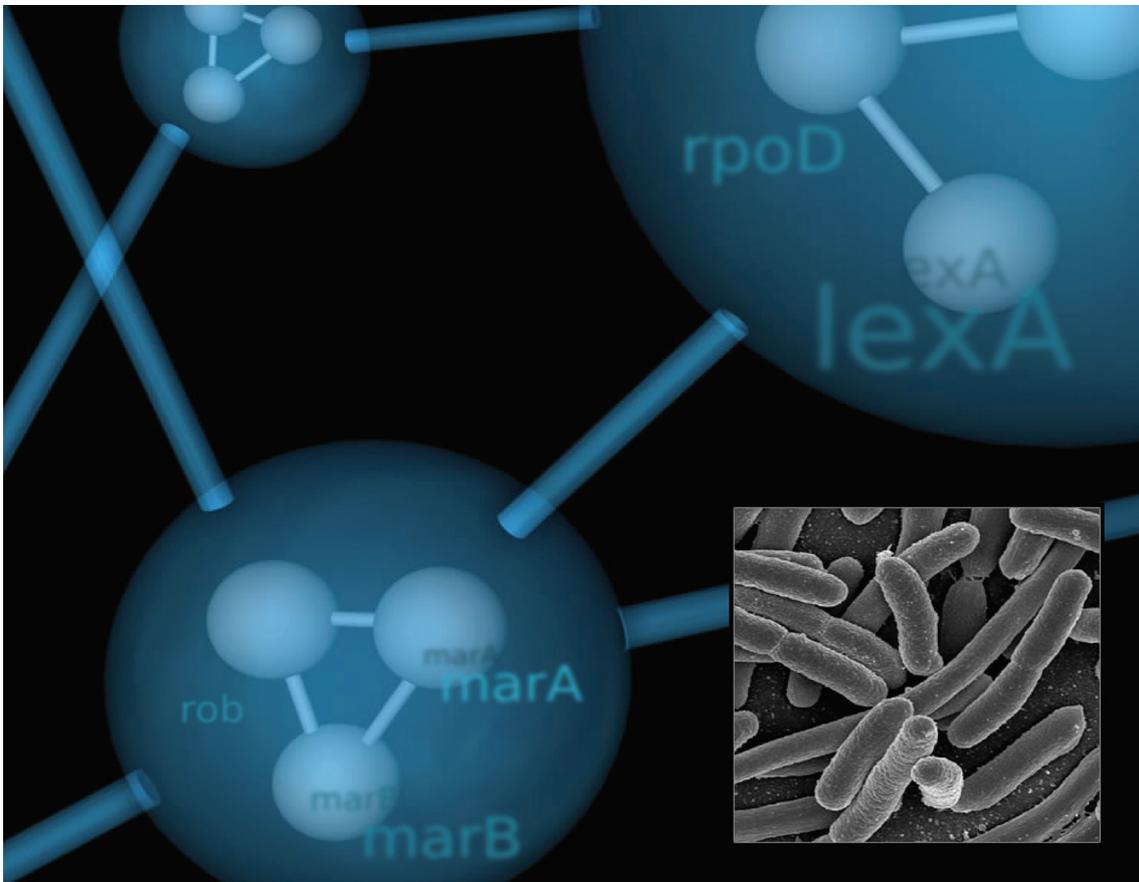


Molecular BioSystems

This article was published as part of the

Computational and Systems Biology
themed issue

Please take a look at the full [table of contents](#) to access the
other papers in this issue.



Transcription regulatory networks in *Caenorhabditis elegans* inferred through reverse-engineering of gene expression profiles constitute biological hypotheses for metazoan development†

Vanessa Vermeirssen,^{ab} Anagha Joshi,^{ab} Tom Michoel,^{ab} Eric Bonnet,^{ab}
Tine Casneuf^{ab} and Yves Van de Peer^{*ab}

Received 23rd April 2009, Accepted 12th June 2009

First published as an Advance Article on the web 17th July 2009

DOI: 10.1039/b908108a

Differential gene expression governs the development, function and pathology of multicellular organisms. Transcription regulatory networks study differential gene expression at a systems level by mapping the interactions between regulatory proteins and target genes. While microarray transcription profiles are the most abundant data for gene expression, it remains challenging to correctly infer the underlying transcription regulatory networks. The reverse-engineering algorithm LeMoNe (learning module networks) uses gene expression profiles to extract ensemble transcription regulatory networks of coexpression modules and their prioritized regulators. Here we apply LeMoNe to a compendium of microarray studies of the worm *Caenorhabditis elegans*. We obtain 248 modules with a regulation program for 5020 genes and 426 regulators and a total of 24 012 predicted transcription regulatory interactions. Through GO enrichment analysis, comparison with the gene–gene association network WormNet and integration of other biological data, we show that LeMoNe identifies functionally coherent coexpression modules and prioritizes regulators that relate to similar biological processes as the module genes. Furthermore, we can predict new functional relationships for uncharacterized genes and regulators. Based on modules involved in molting, meiosis and oogenesis, ciliated sensory neurons and mitochondrial metabolism, we illustrate the value of LeMoNe as a biological hypothesis generator for differential gene expression in greater detail. In conclusion, through reverse-engineering of *C. elegans* expression data, we obtained transcription regulatory networks that can provide further insight into metazoan development.

Introduction

The availability of genome sequences for many organisms has resulted in the development of genomic technologies, which enable the study of molecular biology at a systems biology level. Systems biology aims to obtain an integrative view of biological processes by characterizing complex interactions in an organism. While it remains experimentally challenging to improve on high-throughput technologies for the mapping of different biological molecules and their interactions, the extraction of meaningful biological hypotheses out of the numerous data continues to be the computational bottleneck.

To date, gene expression profiles are one of the most abundant high-throughput sources of data. Several reverse-engineering methods have been developed to extract transcription hypotheses from microarray data.^{1–3} They model the data in transcription regulatory networks that describe gene expression at a systems level as a function of regulatory inputs specified by interactions between regulatory proteins and DNA. The

basic assumption is that regulatory proteins are themselves regulated by transcription, so that their expression profiles provide information about their activity level.

Differential gene expression is an important driving force in the development, function and pathology of eukaryotic organisms. Most regulation of gene expression occurs at the level of transcription, where specific transcription factors bind DNA in order to activate or repress the expression of a gene. Transcription is further influenced by protein–protein interactions between transcription factors themselves or between transcription factors and cofactors, chromatin modifying factors or the basal transcription machinery.⁴ In addition, signal transduction pathways can target all these regulatory proteins, switching them “on” and “off” through phosphorylation by kinases or dephosphorylation by phosphatases.⁵ Hence, inferring transcription regulatory networks through reverse-engineering of gene expression profiles should lead to a better understanding of development in health and disease.

However, the noise inherent to high-throughput data, the imbalance between the number of genes (variables) and the number of experiments (data) and the fact that gene expression profiles measure indirect regulator–gene interactions, all pose difficulty for inferring biologically relevant transcription regulatory interactions. In order to anticipate these complexities in an intelligent way, we developed the LeMoNe (learning

^a Department of Plant Systems Biology, VIB, B-9052 Ghent, Belgium

^b Bioinformatics and Evolutionary Genomics,

Department of Plant Biotechnology and Genetics, Ghent University, B-9052 Ghent, Belgium. E-mail: yves.vandeppeer@psb.vib-ugent.be

† This article is part of a *Molecular BioSystems* themed issue on Computational and Systems Biology.

module networks) algorithm.^{6,7} LeMoNe is a probabilistic module networks framework, like Genomica,³ in which coregulated genes share the same parents and conditional distributions in a Bayesian network, hence limiting the number of variables, reducing the complexity of the learning and leading to more robust solutions. Using Gibbs sampling, LeMoNe groups genes in coexpression modules and then using a fuzzy decision tree, it assigns regulators to these gene modules, based on how well the regulators explain the condition-dependent expression behavior of the module. To filter out noise, LeMoNe extracts an ensemble solution from many equiprobable solutions,^{6,8} both in the partitioning of genes in modules and in the assignment of regulators. Moreover, the algorithm ranks model predictions by providing weights to the assigned regulators. Using benchmark expression data we have shown that LeMoNe successfully prioritizes known transcription regulatory interactions, reaching positive predictive values of more than 90% in *Escherichia coli* and more than 40% in *Saccharomyces cerevisiae*,⁹ and that it significantly outperforms the original module network framework Genomica.⁶ Providing the algorithm with a list of candidate regulators that not only includes transcription factors, but also chromatin modifiers, signal transducers and other regulatory proteins, we take the indirect nature of inferred interactions from gene expression profiles into account.

In this study, we infer transcription regulatory networks for the worm *Caenorhabditis elegans* through LeMoNe. *C. elegans* is one of the most widely used model organisms, due to its biological simplicity and the advantages it offers in terms of its transparent body, invariable cell number, small genome size, rapid life cycle, mode of reproduction and ease of maintenance.¹⁰ Worm research has already led to important discoveries in diverse fields from development, signal transduction, cell death and ageing to RNAi. While a genome-wide collection of transcription regulatory interactions has been reported for *E. coli* and *S. cerevisiae*,^{11,12} transcription regulatory networks for metazoan organisms, including *C. elegans*, have been mapped experimentally only to a limited extent. In *C. elegans*, direct transcription factor regulatory interaction networks have been identified through chromatin immunoprecipitation,^{13,14} and yeast one-hybrid experiments,^{15–17} while indirect regulatory interaction networks have mainly been identified through combinations of gene knockout or knockdown and gene reporter, RT-PCR or microarray experiments.^{18–21} In the latter case, *cis*-regulatory motif finding algorithms have been applied in order to pinpoint the transcription factors directly responsible for the change in gene expression.²² Reverse-engineering could significantly increase the knowledge on transcription regulatory networks in *C. elegans* and put forward a set of interactions worth the experimental time, cost and effort to validate them *in vivo*. To our knowledge, the literature reports one probabilistic model called TRANSMODIS, which integrates *cis*-regulatory sequence and transcription factor perturbation expression information in order to identify direct targets of the transcription factor DAF-16 in *C. elegans*.²³ Since the consensus binding motif is only known for a few transcription factors and perturbation microarray experiments are limited, this algorithm is not yet applicable on a genome-wide scale.

We applied LeMoNe to a carefully preprocessed compendium of *C. elegans* Affymetrix gene expression profiles. We verified the coexpression modules for functional coherence by GO enrichment and the presence of gene–gene associations from WormNet.²⁴ WormNet is a probabilistic functional gene network of protein-encoding genes of *C. elegans*, constructed using a modified Bayesian integration of many different data types from several different organisms, with each data type weighted according to how well it links genes that are known to function together in *C. elegans*. Next, we compared the predicted transcription regulatory interactions to a set of known regulatory interactions in *C. elegans*,^{15,17,25,26} and to the predicted interactions obtained by another reverse-engineering method, the context likelihood of relatedness (CLR) algorithm.² CLR is a direct reverse-engineering method that scores all possible pairwise regulator–target gene interactions based on the mutual information of their expression profiles as compared to an interaction specific background distribution. Recently, we have observed that module-based and direct reverse-engineering methods retrieve distinct parts of underlying transcription regulatory networks.⁹ Furthermore, we characterize some transcription regulatory networks that are put forward by coexpression modules and their regulators for which biologically functional information is available. Overall, we demonstrate that LeMoNe is highly qualified to generate biologically relevant hypotheses on transcription regulatory networks in *C. elegans*.

Results and discussion

LeMoNe predicts functionally coherent coexpression modules

From the Affymetrix microarray compendium of 11 713 genes and 155 conditions, LeMoNe consistently partitioned 5020 genes in 248 coexpression modules of at least three genes and a regulation program. These modules contained on average 20 genes, ranging from three up to 203 genes. An overview of the LeMoNe workflow and post-processing is depicted in Fig. 1. Table 1 lists all modules for which there was a GO enrichment and/or overlap with known and/or CLR predicted regulatory interactions (see below).‡

In order to assess the true biological nature of the modules, we looked into the functional coherence of the modules. Coexpressed and coregulated genes are more likely to function in a similar biological process. First, we looked at the gene ontology and found 72 modules with a significant GO biological process enrichment. The actual number of functional coherent modules may be considerably higher, since only about 40% of *C. elegans* genes are annotated with a biological process in WormBase (<http://www.wormbase.org>). In addition, 41 modules contained only three or four genes and those received no GO enrichment, also because we required the significant GO enriched biological process term to be attributed to at least three genes in the module.

‡ The LeMoNe predicted modules with their genes and regulators are available through a web interface at http://bioinformatics.psb.ugent.be/supplementary_data/vamei/Celegans_lemone_clr. CLR predictions for the genes in the modules can also be viewed.

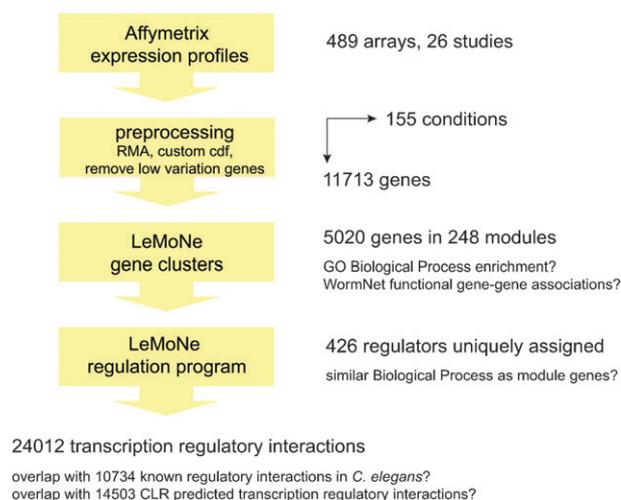


Fig. 1 Overview of the reverse-engineering of gene expression profiles for *C. elegans*.

As a second, independent measure of functional coherence, we evaluated whether the genes in a module shared functional gene–gene associations reported in WormNet.²⁴ WormNet integrates diverse datasets from different organisms in a gene–gene network, where tightly linked genes are very likely to be involved in the same biological process. WormNet comprises 384 700 linkages among 16 113 genes. Worm cDNA coexpression data, human interologs and functional associations of yeast orthologs are the main data contributions to WormNet. WormNet’s coexpression data are all obtained from the Stanford Microarray Database, hence from cDNA microarrays. In total, 161 LeMoNe modules contained genes that are also connected to one another in WormNet, with 84 modules having 50% or more of the genes being linked in WormNet, 37 of which also displayed a significant GO enrichment.

When looking into detailed gene information for some of the modules that did not have a significant GO enrichment, we nevertheless observed that the genes are related to similar biological processes (*e.g.* module 122 genes are active in the germline and early embryogenesis; module 108 genes function in ciliated sensory neurons, see below). Moreover, most modules contain uncharacterized genes, for which based on the “guilt by association” principle, we can now predict that they function in a specific biological process. Hence, we can conclude that LeMoNe correctly identified modules of genes that are coexpressed over at least some of the conditions and that function together in a specific biological process.

LeMoNe prioritizes biologically relevant regulators

For the assignment of regulators, LeMoNe used as an input 1063 candidate regulators with expression information in the compendium. These were mostly worm transcription factors (855),²⁶ but also included proteins involved in signal transduction (187) and chromatin modification (6) and proteins that have been shown to bind DNA in yeast one-hybrid assays (15).^{15,17} Each of the 248 coexpression modules received on average four predicted regulators with a weight of 100 or higher; only one module had none and the maximum number of regulators

per module was eight. In total, 426 regulators were uniquely assigned to coexpression modules, 49% of which were only assigned to one module and less than one percent was assigned to 10 modules or more. By relating all predicted regulators of a module to all genes in that module, we ended up with 24 012 LeMoNe predicted interactions.

Next, we assessed to what extent a biologically relevant regulation program or candidate regulator is inferred for a module. We have previously shown that the score of a regulator efficiently prioritizes interactions that have been observed experimentally,⁹ and therefore high-scoring regulators are more likely to be true regulators. This justifies the restriction to those regulators having a weight of 100 or higher, which is less than 1% of all regulators LeMoNe initially had assigned. The top regulator of a module has on average a weight of 553. A transcription regulatory interaction is also more plausible, if the regulator is involved in a similar biological process as the genes in the module. We could only investigate this for modules and regulators that were sufficiently characterized (see below).

In addition, we compared the predicted transcription regulatory interactions from LeMoNe with a set of 1120 known regulatory interactions in *C. elegans*. These data consisted of 598 yeast one-hybrid protein–DNA interactions,^{15,17,26} and 525 Wormbase WS195 regulatory interactions. The first are direct regulator–target binding interactions, while for the latter this is not necessarily the case. We found seven regulator–gene interactions overlapping between both sets, respectively, in modules 21, 29 and 34 (Table 2). Due to the limited coverage of reported regulatory interactions, this low number is not surprising. Modules 21, 29 and 34 are all related to embryonic development, meiosis and oogenesis. In the hermaphrodite *C. elegans*, oocyte maturation, ovulation, fertilization and initiation of embryogenesis are tightly coordinated.²⁷ These modules also share regulators like the homologs MEX-5 and MEX-6 and the homologs OMA-1 and OMA-2, which are involved in oogenesis and/or early embryogenesis, indicating that these modules might be coregulated (WormBase).^{28,29} Module 34 contains most known regulatory interactions and is discussed in more detail below. In addition, we observed a known regulatory interaction between genes in module 141. This could signify that SEX-1, which is present as a gene in the module, is actually regulating module 141. There is indeed extra evidence for this hypothesis, since in addition to SEX-1 and DPY-27, proteins functioning in sex determination and oviposition, this module contains ALY-1, which is also involved in sex determination, and R07E5.3, which is also involved in oviposition (WormBase). Finally, three known regulatory interactions were found between regulators only, for modules 56, 115 and 167. These point to regulatory chains that are governing the regulatory program of a module.

When looking into the known regulatory interaction list, many genes that share a regulator ended up in the same module, whether or not the regulator is predicted by LeMoNe (see modules 34 and 108, and data not shown). In the case where LeMoNe predicts a different regulator, there might exist a “hidden path” that connects the predicted regulator to the known regulator, since LeMoNe infers indirect regulatory

Table 1 Summary of the modules for which there is a GO enrichment and/or overlap with known and/or CLR predicted regulatory interactions^a

<i>M</i>	# Genes	# Regs	% WormNet	GO enrichment	Known	CLR
1	203	7	65	Sensory perception of chemical stimulus, G-protein coupled receptor protein signaling pathway	—	—
2	83	6	26	Sensory perception of chemical stimulus, G-protein coupled receptor protein signaling pathway	—	—
3	88	6	42	Signal transduction	Yes	—
5	134	6	39	Protein amino acid glycosylation	—	—
6	70	7	50	Sensory perception of chemical stimulus	—	—
7	57	5	38	Sensory perception of chemical stimulus	Yes	—
8	52	4	5	G-protein coupled receptor protein signaling pathway	—	—
9	57	2	0	—	Yes	—
10	64	6	31	Negative regulation of vulval development	—	—
11	44	3	47	Phosphate metabolic process, post-translational protein modification, protein amino acid (de)phosphorylation	—	—
12	50	2	30	—	Yes	—
13	41	4	0	—	Yes	—
15	49	5	10	Response to DNA damage stimulus, meiosis, M phase of meiotic cell cycle, response to stress	—	—
16	93	6	47	Hermaphrodite genitalia development, RNA processing, growth, cytoskeleton organization and biogenesis	—	—
17	66	5	13	G-protein coupled receptor protein signaling pathway, ion transport	Yes	—
19	36	4	10	G-protein coupled receptor protein signaling pathway, locomotory behavior	—	—
20	34	5	80	Molting cycle, collagen and cuticulin-based cuticle	—	Yes
21	47	3	73	Meiosis, M phase of meiotic cell cycle	Yes	—
22	38	7	56	Embryonic development, growth, larval development, translation	Yes	—
23	58	4	17	Defecation, excretion, secretion	—	—
26	145	5	60	—	—	Yes
27	38	7	75	Proteolysis	—	—
28	34	4	100	Translation, growth, larval development, embryonic development, reproduction	—	—
29	40	6	89	Embryonic development	Yes	—
34	33	8	97	Meiosis, M phase of meiotic cell cycle, oogenesis, meiotic chromosome segregation	Yes	—
35	38	5	6	G-protein coupled receptor protein signaling pathway	—	—
36	65	6	42	Fatty acid metabolic process, monocarboxylic acid metabolic process	—	—
40	26	5	72	Protein amino acid dephosphorylation	—	—
41	36	6	40	Hermaphrodite genitalia development, sex differentiation, gastrulation with mouth forming first, growth	—	—
44	23	7	32	—	Yes	—
47	25	4	42	Translation	—	—
49	33	3	7	—	Yes	—
50	21	6	24	Regulation of transcription, DNA-dependent	—	—
52	28	4	0	—	Yes	—
53	48	4	63	Phosphate metabolic process, post-translational protein modification, protein amino acid (de)phosphorylation	—	—
55	23	3	18	—	Yes	—
56	20	5	0	—	Yes	—
57	26	6	21	Signal transduction	—	—
59	27	4	22	—	Yes	—
62	26	7	46	Embryonic development, growth	—	—
63	17	5	56	Oogenesis, hermaphrodite genitalia development, sexual differentiation	—	—
64	21	6	95	Locomotion, molting cycle, collagen and cuticulin-based cuticle	Yes	—
65	30	8	9	G-protein coupled receptor protein signaling pathway	Yes	—
67	19	3	88	Embryonic development, larval development, growth, biopolymer catabolic process	—	—
68	24	6	35	—	Yes	—
71	22	4	55	Macromolecule biosynthetic process	—	—
75	15	4	29	Post-translational protein modification, phosphate metabolic process	—	—
77	20	5	35	—	Yes	—
79	20	3	25	Vesicle-mediated transport	—	—
81	15	5	50	Transcription, growth, hermaphrodite genitalia development	—	—
83	15	6	0	—	Yes	—
86	18	4	100	Translation, growth, larval development, embryonic development, reproduction	—	—
88	34	5	22	Aromatic compound metabolic process	Yes	—
89	20	5	33	Regulation of transcription, DNA-dependent	Yes	—
91	14	6	92	Molting cycle, collagen and cuticulin-based cuticle, locomotion, body morphogenesis	—	—
92	15	1	73	Larval development, growth, embryonic development, reproduction	—	—
95	19	5	12	Embryonic development	—	—
96	19	3	12	Neurological system process, ion transport	—	—

Table 1 (continued)

<i>M</i>	# Genes	# Regs	% WormNet	GO enrichment	Known	CLR
97	21	3	95	Proteolysis, cell death	—	—
100	21	4	0	—	—	Yes
101	39	4	63	Growth, larval development, electron transport chain	—	—
105	40	4	36	Meiotic chromosome segregation, M phase of meiotic cell cycle	—	—
106	16	4	44	Metabolic process	—	—
108	15	7	0	—	Yes	—
109	14	3	0	Growth, locomotion	—	—
113	19	4	47	Amino acid and derivative metabolic process	—	—
115	11	4	0	—	Yes	—
117	14	4	0	—	Yes	—
124	12	3	75	Cytoskeleton organization, cell cycle, growth	—	—
125	26	3	17	Regulation of transcription, DNA-dependent	—	—
126	15	5	93	Translation, growth, larval development	—	—
128	22	2	24	Energy derivation by oxidation of organic compounds, generation of precursor metabolites and energy, growth	—	—
130	17	3	12	Determination of adult life span, aging	—	—
133	12	2	56	Molting cycle, collagen and cuticulin-based cuticle	—	—
135	14	5	62	Nitrogen compound metabolic process, carboxylic acid metabolic process, lipid metabolic process	—	—
137	14	3	54	Generation of precursor metabolites and energy, oxidative phosphorylation, growth	—	—
140	11	5	89	Ubiquitin-dependent protein catabolic process, proteolysis	—	—
141	11	7	50	Oviposition	Yes	—
142	9	2	67	Generation of precursor metabolites and energy, positive regulation of growth, ion transport	—	—
148	12	2	100	Translation, growth, larval development, embryonic development, reproduction	—	—
150	22	3	33	Metabolic process	—	—
152	22	4	55	Regulation of vulval development, growth, larval development, embryonic development	—	—
161	6	4	100	Translation, growth, larval development, embryonic development, reproduction	—	—
167	7	3	0	—	Yes	—
171	6	3	80	Locomotion	—	—
172	9	5	100	Cytokinesis, hermaphrodite genitalia development, sex differentiation	—	—
188	5	2	80	Nucleosome assembly, chromatin assembly or disassembly, chromosome organization and biogenesis	—	—
203	7	4	57	Growth	—	—
221	5	5	50	Anatomical structure development	—	—
265	5	2	75	Regulation of transcription, DNA-dependent	—	—
288	3	2	67	—	—	Yes
315	5	5	0	—	Yes	—

^a The module number (*M*), the number of genes (# Genes), the number of regulators (# Regs), the percentage of genes that are connected with one another in WormNet (% WormNet), the most significant GO biological process terms (GO enrichment), overlap with known or derived regulatory interactions (Known) and overlap with CLR predictions (CLR) are given. The modules that are discussed in more detail are in bold.

Table 2 Overlap between LeMoNe predictions and reported regulatory interactions in *C. elegans* (either yeast one-hybrid protein–DNA interactions (PDI) or WormBase WS195 regulatory interactions (RI))^a

Type	<i>M</i>	Regulator	Target	Type	Regulator in module?	<i>W1</i>	<i>W2</i>	<i>R1</i>	<i>R2</i>
REG–GENE	21	MEX-6	CDC-25.1	RI	No	630	—	2	—
REG–GENE	29	OMA-2	MEX-1	RI	No	349	—	3	—
REG–GENE	34	OMA-1	PGL-1	RI	Yes	161	—	6	—
REG–GENE	34	OMA-1	MEX-5	RI	Yes	161	—	6	—
REG–GENE	34	OMA-1	PIE-1	RI	Yes	161	—	6	—
REG–GENE	34	OMA-2	PGL-1	RI	Yes	312	—	3	—
REG–GENE	34	OMA-2	PIE-1	RI	Yes	312	—	3	—
GENE–GENE	141	SEX-1	DPY-27	RI	Yes	—	—	—	—
REG–REG	56	FOS-1	EGL-43	RI	No	262	104	2	5
REG–REG	115	CEH-8	UNC-42	PDI	No	130	576	4	1
REG–REG	167	OMA-1	POS-1	RI	No	122	229	3	1

^a The regulatory interaction was either between a predicted regulator and a module gene (REG–GENE), between two genes in the module, one of them being a regulator (GENE–GENE), or between two predicted regulators, one of them regulating the other (REG–REG). *M* = module number; *W1*, *R1* are, respectively, weight and rank of regulator, if regulator was predicted as regulator by LeMoNe; *W2*, *R2* are, respectively, weight and rank of target, if target was predicted as regulator by LeMoNe.

interactions and this does not necessarily implicate direct binding between regulator and target gene. Therefore, we decided to do a more thorough comparison by calculating all possible indirect paths between regulators and target genes from the reported set of *C. elegans* regulatory interactions. We retrieved 9614 derived regulatory interactions (see Methods). Comparing this dataset to the LeMoNe predictions revealed an additional overlap of 28 regulator–gene, five gene–gene and eight regulator–regulator interactions (Table 3). Based on this, we could infer a whole regulatory path for a module and relate more regulators to the module genes. For example, for module 49, two possible regulatory paths between the regulator LIN-54 and the target gene ZAG-1 are one involving the regulators ZTF-4, RNT-1 and TLP-1, and the other involving the regulators TTX-1 and ODR-7. We characterize the regulatory paths for module 108 in more detail below. There might be several explanations why LeMoNe did not predict

the regulators of the hidden path, *e.g.* they might have low expression variation over the conditions and diverse biological functions or they might be regulated at the post-transcriptional level. Overall, many modules contained known or derived regulatory interactions (Table 1).

Comparison with CLR, another reverse-engineering method

We also predicted transcription regulatory interactions for *C. elegans* on the Affymetrix microarray compendium by means of another reverse-engineering method, CLR.² CLR prioritizes regulator–gene interactions using a *z*-score. With a *z*-score cut-off of five (p -value $< 3 \times 10^{-7}$), CLR inferred 14 503 transcription regulatory interactions for *C. elegans*, between 4048 genes and 583 regulators. We did not find any overlap between CLR and the set of known regulatory interactions.

We have previously observed that CLR and LeMoNe infer mainly different parts of the underlying transcription

Table 3 Overlap between LeMoNe predictions and derived interactions from reported regulatory interactions in *C. elegans* (see Methods)^a

Type	<i>M</i>	Regulator	Target	Regulator in module?	<i>W</i> 1	<i>W</i> 2	<i>R</i> 1	<i>R</i> 2
REG–GENE	7	TBX-35	EXP-1	No	526	—	2	—
REG–GENE	9	SEM-4	LIN-11	No	210	—	2	—
REG–GENE	12	SEM-4	NRX-1	Yes	156	—	2	—
REG–GENE	13	SEM-4	MAB-9	No	112	—	4	—
REG–GENE	17	ZTF-8	GCY-22	No	107	—	5	—
REG–GENE	22	UNC-120	XBX-6	No	120	—	6	—
REG–GENE	22	ZTF-8	XBX-6	No	149	—	4	—
REG–GENE	49	LIN-54	ZAG-1	No	362	—	2	—
REG–GENE	52	ALR-1	T22C8.3	No	206	—	2	—
REG–GENE	55	SEM-4	EFF-1	No	1044	—	1	—
REG–GENE	57	ALR-1	SYD-2	No	175	—	5	—
REG–GENE	57	CEH-20	SYD-2	No	237	—	2	—
REG–GENE	57	AST-1	SYD-2	No	205	—	4	—
REG–GENE	59	TAG-233	DAF-19	Yes	113	—	4	—
REG–GENE	65	ZTF-2	CHE-12	No	144	—	6	—
REG–GENE	65	UNC-42	CHE-12	No	247	—	3	—
REG–GENE	68	CEH-20	LIT-1	No	436	—	2	—
REG–GENE	77	DMD-4	SOD-3	No	149	—	3	—
REG–GENE	88	ZC204.12	MDL-1	No	123	—	3	—
REG–GENE	89	TTX-1	ZIG-4	Yes	162	—	4	—
REG–GENE	89	AST-1	ZIG-4	No	210	—	3	—
REG–GENE	108	LIN-11	BBS-5	No	145	—	4	—
REG–GENE	108	UNC-42	BBS-5	No	388	—	1	—
REG–GENE	108	LIN-11	DYF-11	No	145	—	4	—
REG–GENE	108	UNC-42	DYF-11	No	388	—	1	—
REG–GENE	117	UNC-42	CEH-2	No	465	—	2	—
REG–GENE	117	UNC-42	CHE-11	No	465	—	2	—
REG–GENE	167	OMA-1	NHR-79	No	122	—	3	—
GENE–GENE	44	ZTF-19	PHA-4	Yes	—	—	—	—
GENE–GENE	59	ZTF-12	DAF-19	Yes	—	—	—	—
GENE–GENE	89	LIN-39	SYG-2	Yes	—	—	—	—
GENE–GENE	89	LIN-39	ZIG-4	Yes	—	—	—	—
GENE–GENE	89	MAB-5	ZIG-4	Yes	—	—	—	—
REG–REG	3	TBX-35	CHE-1	No	276	294	4	3
REG–REG	52	ALR-1	AST-1	No	206	169	2	4
REG–REG	57	ALR-1	AST-1	No	175	205	5	4
REG–REG	57	CEH-20	AST-1	No	237	205	2	4
REG–REG	64	SEM-4	UNC-86	No	131	105	4	5
REG–REG	83	ZTF-1	NHR-68	No	139	104	3	6
REG–REG	108	LIN-11	UNC-42	No	145	388	4	1
REG–REG	315	AST-1	UNC-30	No	308	108	2	5

^a The regulatory interaction was either between a predicted regulator and a module gene (REG–GENE), between two genes in the module, one of them being a regulator (GENE–GENE), or between two predicted regulators, one of them regulating the other (REG–REG). *M* = module number; *W*1, *R*1 are, respectively, weight and rank of regulator, if regulator was predicted as regulator by LeMoNe; *W*2, *R*2 are, respectively, weight and rank of target, if target was predicted as regulator by LeMoNe.

Table 4 Overlap between LeMoNe predictions and CLR predictions in *C. elegans*^a

<i>M</i>	Regulator	Target	Regulator in module?	Weight	Rank	<i>z</i> -score
20	C12D12.5	T03G6.1	No	392	2	6.17
20	C12D12.5	K04H4.2	No	392	2	6.07
20	C12D12.5	BE10.2	No	392	2	6.04
20	C12D12.5	K10D3.4	No	392	2	5.48
20	C12D12.5	C15C6.1	No	392	2	5.36
20	C12D12.5	Y64H9A.2	No	392	2	5.10
20	C12D12.5	W04G3.1	No	392	2	5.06
20	C12D12.5	T19A5.3	No	392	2	5.03
20	C12D12.5	C14A4.9	No	392	2	5.01
26	C04F5.9	C32D5.11	No	131	5	5.37
100	2L52.1	Y95B8A.8	Yes	658	1	5.91
100	2L52.1	Y39G10AR.11	Yes	658	1	5.60
100	2L52.1	Y71G12B.13	Yes	658	1	5.42
100	2L52.1	Y53F4B.13	Yes	658	1	5.01
288	M18.8	F48C1.5	No	405	1	5.36

^a All are regulator–gene interactions. *M* = module number.

regulatory networks of *E. coli* and *S. cerevisiae*.⁹ While CLR makes predictions for a higher number of regulators (more target gene hubs), LeMoNe predicts more target genes for fewer regulators (more regulator hubs). The advantage of LeMoNe over CLR is that a predicted transcription regulatory interaction can be directly related to a biological process through the module context. While they are highly complementary, this also indicates that interactions predicted by both methods are of great value.⁹ For the genes in LeMoNe modules, CLR predicted 6602 interactions.‡ When comparing CLR to LeMoNe, we observed 15 predicted regulator–gene interactions in common (Table 4). Especially for module 20 and module 100, CLR predicted interactions between many genes in the module and, respectively, the second-best and best regulator of that module. Module 20 is discussed in more detail below. In addition, we observed several modules that illustrate the complementarity of LeMoNe and CLR (see modules 108 and 142 below).

We discuss four modules in greater detail: module 20, for which we observed a high overlap between LeMoNe and CLR (Table 4); module 34, for which we retrieved many reported regulatory interactions (Table 2); module 108, for which we found reported hidden regulatory paths (Table 3) and module 142. For all these modules external evidence further supports the fact that the predicted regulator(s) and the module genes function in a similar biological process.

Molting and larval development

Module 20 (34 genes) contains several genes involved in molting (ACN-1, BUS-8, PTR-1 and T19A5.3), which is the process of shedding of *C. elegans*' extracellular matrix, the cuticle, at each larval stage (Fig. 2). PTR-1 is a sterol-sensing domain protein. In addition, the module contains BE10.2, belonging to the KOG group of sterol C5 desaturase (NCBI) and involved in fatty acid biosynthesis; ELO-4, a polyunsaturated fatty acid elongase; and CLEC-180, a C-type lectin belonging to the KOG group of low-density lipoprotein receptors. This is in agreement with the fact that most likely a steroid hormone triggers molting, since this process requires cholesterol, the biosynthetic precursor of all steroid

hormones.³⁰ Many genes in module 20 have not yet been characterized, but based on their association with the genes above for which functional information is available, we can hypothesize that they also function in molting. All genes in the module are downregulated in embryos and adults, while they are upregulated during larval development, in accordance with the timing of molting. C12D12.5, which is predicted as a regulator for nine genes in this module by CLR and as the second-best regulator for this module by LeMoNe, is an uncharacterized HMG transcription factor (WormBase). From the predictions of LeMoNe and CLR, we can postulate that this transcription factor functions during larval development and is implicated in the molting cycle. Module 20 has several other regulators, which also might be active in these processes (WormBase). The top regulator RAB-3 is a RAS small GTPase signal transducer, which is expressed in all neurons and connected to many biological processes, especially synaptic transmission, chemotaxis, locomotion, pharyngeal pumping and mating behavior. The third regulator ZBED-6 is required for larval development and growth and is involved in the molting cycle. The fourth regulator F17A9.3 is reported to be expressed during larval development. For the fifth regulator NHR-230 there is not much biological information available and its expression profiles do not correspond so well with those of the module genes. Interestingly, we found some of these regulators back in other modules that were GO enriched in the molting cycle. ZBED-6 was predicted as the second regulator both in module 64 and module 133. Members of the RAS small GTPase family were also predicted as regulators in module 91 and module 133. Finally, F17A9.3 was the best regulator for module 91. This again suggests some coregulation for modules involved in a similar biological process.

Meiosis and oogenesis

Module 34 (33 genes) is GO enriched for the M phase of the meiotic cell cycle, cell division, oogenesis and cell differentiation, in addition to embryonic development (Fig. 3). In the hermaphrodite *C. elegans*, the germline produces both male and female gametes, sperm and oocytes, respectively. While oocytes are produced throughout adult life, sperm is generated

Module 20

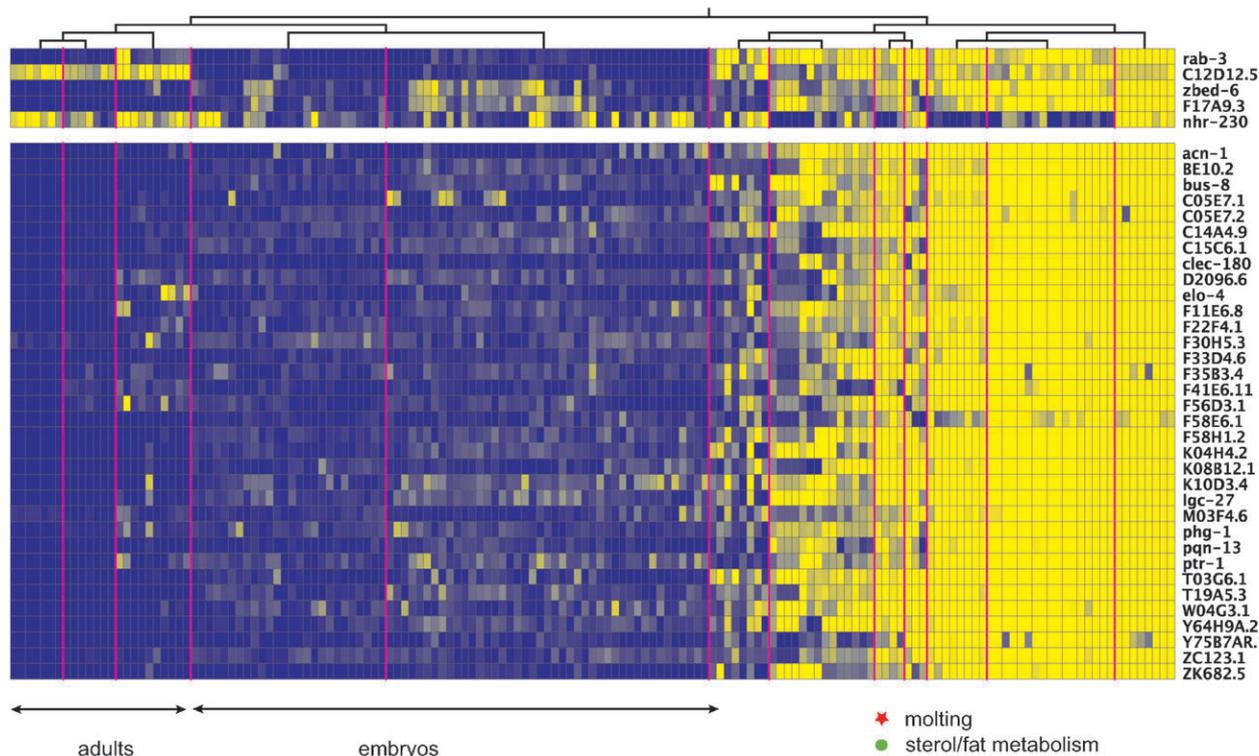


Fig. 2 Predicted regulators and module genes for module 20, which is involved in molting. Red vertical lines partition the different condition clusters. Yellow: upregulated genes; blue: downregulated genes.

during the last larval stage. The adult germline exhibits distal–proximal polarity with a mitotic cell population located at the most distal end of the gonad, meiotic cells extending proximally and formation of sperm and oocytes occurring in the proximal part of the gonad arm. This is followed, both in time and space, by oocyte maturation, ovulation, fertilization and initiation of embryogenesis (WormAtlas, www.wormatlas.org). Hence, meiosis, oogenesis and embryogenesis are tightly coupled processes in worms.²⁷ The majority of the genes in this module are expressed in the medial germline, where meiosis takes place (WormBase) (Fig. 3). In addition, most genes were reported to be “oogenesis-enriched” according to a *C. elegans* genome-wide analysis of gene expression to identify germline- and sex-regulated genes.³¹ Genes with high expression during oogenesis are either genes required for oocyte differentiation or maternal mRNAs required for early embryogenesis. Several genes that encode maternal mRNAs are present in the module: *mex-5*, *mex-6*, *pie-1*, *cey-3*, *oma-1* and *oma-2*. They appear to be under translational control and play a significant role in establishing soma–germline asymmetry in the early embryo.²⁸ PUF domain RNA binding proteins like PUF-5, which are also in the module, have been reported to control maternal mRNAs during late oogenesis.³² This module contains the highest number of regulators and overlapping interactions with *C. elegans* known as regulatory interactions (Table 2). The known regulatory interactions are derived from the observation that *oma-1;oma-2* RNAi embryos have missegregation of PGL-1, PIE-1 and MEX-1 in the early embryo,³³ and that a gain-of-function *oma-1*

mutant has delayed degradation of maternal proteins including SKN-1, PIE-1, MEX-3 and MEX-5.³⁴ From the information we currently have, the most likely regulators of the module are OMA-1 and its paralog OMA-2, which have a reported function in oocyte maturation,³⁵ and ZIM-2, which is required for segregation of chromosomes during meiosis;³⁶ both are processes that are in accordance with the GO enrichment of the module genes.

Ciliated sensory neurons

Despite the absence of a significant GO enrichment, several genes in module 108 are reported to be expressed and function in ciliated sensory neurons (WormBase) (Fig. 4). In addition, most genes and regulators of module 108 are upregulated during neuronal conditions (Fig. 4). In this module we found five derived interactions back from reported *C. elegans* regulatory interactions (Table 3). The proteins BBS-5 and DYF-11 in module 108 are both expressed in all ciliated sensory neurons and share the regulators UNC-42 and LIN-11, which are both transcription factors that function in cell fate determination of sensory neurons (WormBase). LIN-11 is regulating UNC-42 through a path consisting of a regulatory interaction and a protein–DNA interaction involving ODR-7, another transcription factor implicated in sensory neuron cell fate determination (WormBase) (Table 3) (Fig. 5). The expression of BBS-5 and DYF-11 is known to be regulated by DAF-19, an RFX transcription factor that is required for sensory neuron cilium formation (WormBase)

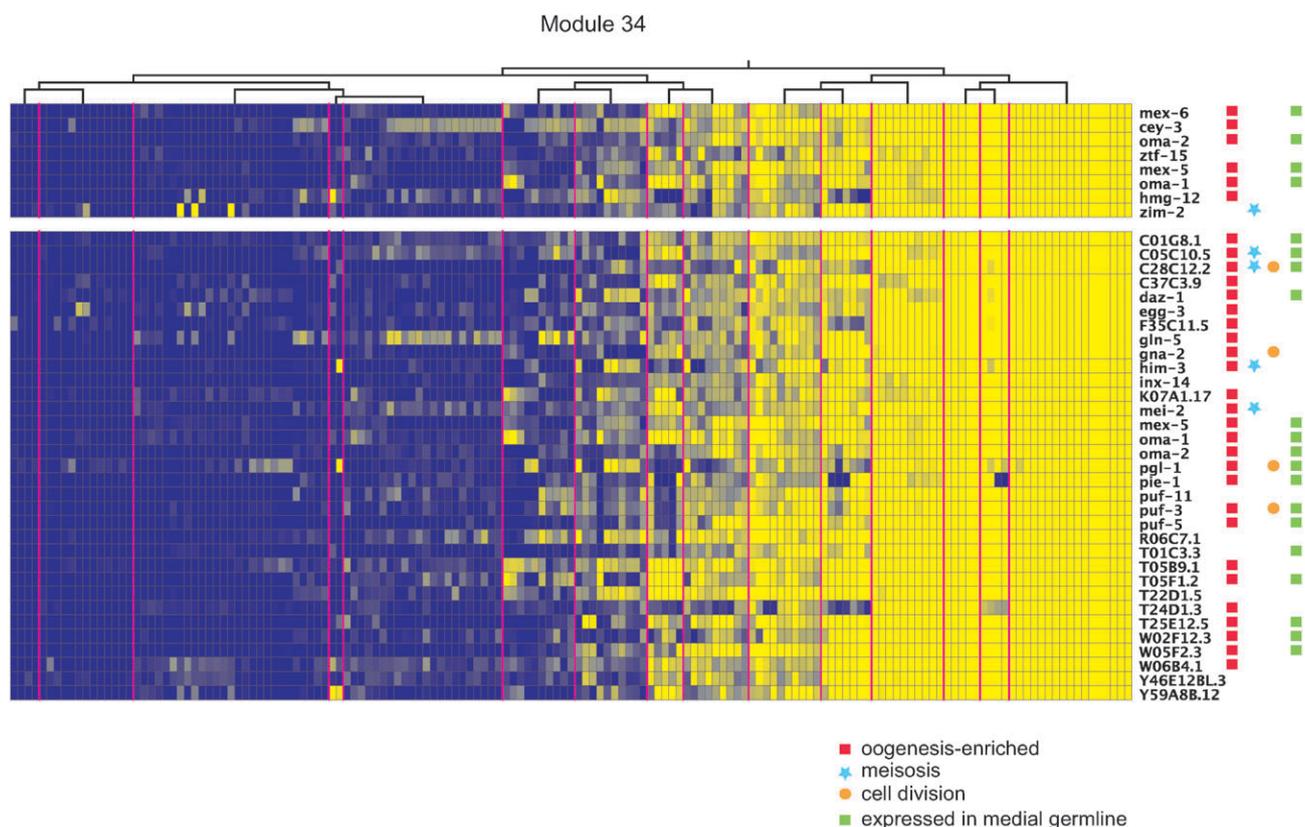


Fig. 3 Predicted regulators and module genes for module 34, which is involved in meiosis and oogenesis. Red vertical lines partition the different condition clusters. Yellow: upregulated genes; blue: downregulated genes.

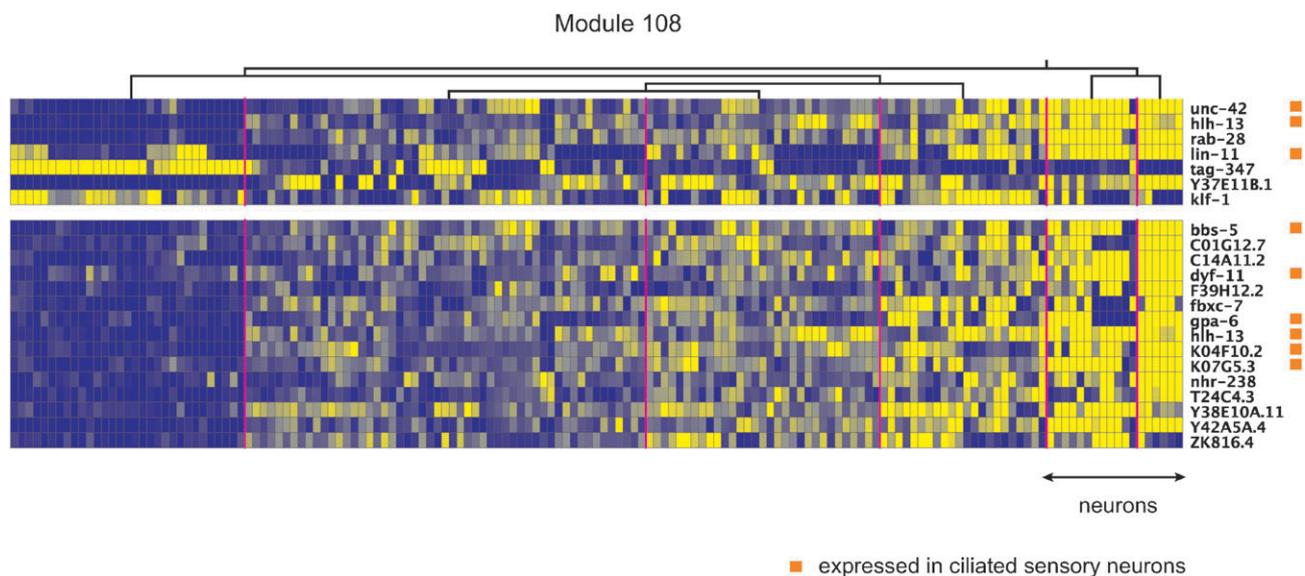


Fig. 4 Predicted regulators and module genes for module 108, which is related to ciliated sensory neurons. Red vertical lines partition the different condition clusters. Yellow: upregulated genes; blue: downregulated genes.

(Fig. 5). Through yeast one-hybrid protein–DNA interaction data we can infer a regulatory path between UNC-42, and hence also LIN-11, and DAF-19 through DAF-3, which is also expressed in sensory neurons (WormBase) (Fig. 5). The characterized paths and the correspondence of the biological function of the transcription factors with that of the module

genes further testify the LeMoNe predictions. CLR predicts ODR-3, a G protein α -subunit that also functions in cilium morphogenesis of sensory neurons, as a regulator for seven out of the 15 genes in module 108, including *bbs-5* and *dyf-11* (Fig. 5). Interestingly, there is a hidden known regulatory path that connects ODR-3 to these two genes (Fig. 5). Although

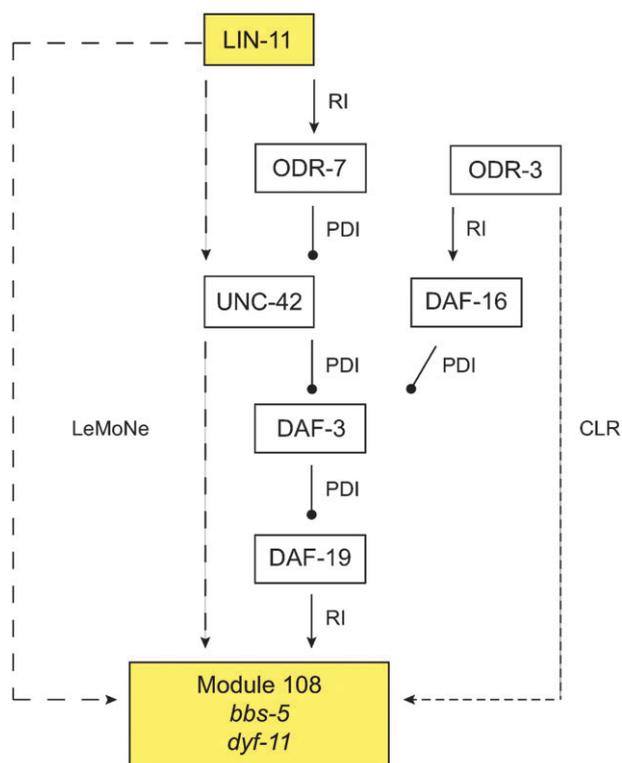


Fig. 5 The transcription regulatory network for module 108, an example where LeMoNe predictions (yellow boxes, dashed lines) are supported by a combination of reported yeast one-hybrid protein–DNA (PDI) and regulatory interactions (RI) (solid lines). UNC-42 and LIN-11 are predicted regulators of module 108. Two genes in this module, *bbs-5* and *dyf-11*, are reported to be regulated by DAF-19 (WormBase), whose promoter is bound by DAF-3 in yeast one-hybrid assays.¹⁷ UNC-42 is bound by ODR-7,¹⁷ which is regulated by LIN-11 (WormBase). CLR predicts ODR-3 as regulator for several genes in module 108 (dashed-dotted line). ODR-3 can also be connected to *bbs-5* and *dyf-11* through a combination of reported interactions.

LeMoNe and CLR predict different regulators for this module, biological functional information and reported regulatory interactions provide evidence for both predictions. This illustrates once more that LeMoNe and CLR can recover

different parts of the underlying transcription regulatory networks.⁹

Energy metabolism, oxidative stress and life span

Module 142 (9 genes) is GO enriched for energy metabolism, growth and ion transport. From the genes in the module we deduced that it is related to the mitochondrial respiratory electron transport chain (Fig. 6). It contains a mitochondrial phosphate carrier protein (F01G4.6), a subunit of cytochrome C oxidase (F26E4.6), two subunits of ATP synthase (F58F12.1 and H28O16.1) and a voltage-gated anion channel in the mitochondrial outer membrane (R05G6.7). Another component involved in respiration is F01F1.12, a fructose biphosphate aldolase, an enzyme in the glycolysis pathway. Energy sources such as glucose are initially metabolized in the cytoplasm, but then products are imported in the mitochondria where catabolism continues to form energy-rich electron donors.³⁷ In the mitochondrial electron transport chain the electrons are transferred to the terminal electron acceptor oxygen *via* a series of redox reactions. These reactions are coupled to the creation of a proton gradient across the mitochondrial inner membrane, which is used to make ATP *via* ATP synthase. Although electron transport occurs with great efficiency, a small percentage of electrons are prematurely leaked to oxygen, resulting in the formation of toxic reactive oxygen species like superoxide and hydrogen peroxide. In this respect, another gene in the module, *prdx-2* encodes a peroxidase that reduces hydrogen peroxide and increases the resistance to oxidative stress.³⁸ Knockout experiments have shown that this gene is also involved in the negative regulation of transcription (WormBase). This provides a relation with the last two genes in this module, R09B3.3, which is involved in RNA processing, and RLA-2, which is a large ribosomal subunit. In addition, these two genes might have yet unknown functional roles in mitochondrial homeostasis.

Moreover, five genes in this module (F01G4.6, H28O16.1, *prdx-2*, R05G6.7, *rla-2*) have an expression pattern in pharynx, intestine and/or body wall muscle, tissues with high metabolic rates that contain many mitochondria (WormBase).³⁷ From the partition of experiments, we observed major differences in gene expression between the conditions of experiment 6 and

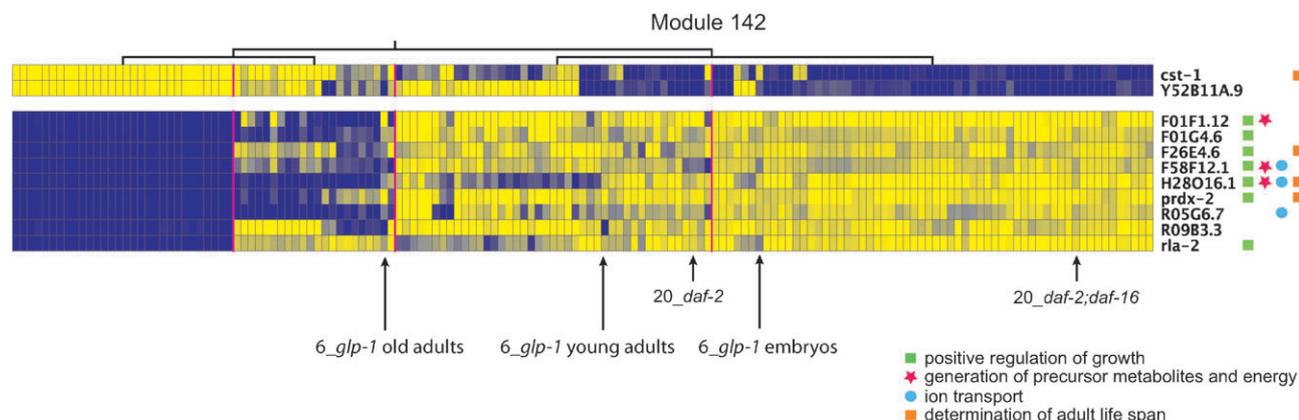


Fig. 6 Predicted regulators and module genes for module 142, which is involved in energy metabolism, growth and ion transport. Red vertical lines partition the different condition clusters. Yellow: upregulated genes; blue: downregulated genes.

experiment 20. In experiment 6, the terminally differentiated *glp-1* mutated embryos, young adults and old adults are present in different experiment clusters, with a higher expression of the module genes in younger animals. This is in accordance with the reported age-related decreased metabolism, mitochondrial function and ATP production.^{39,40} In experiment 20, the module genes are significantly more upregulated in *daf-2;daf-16* mutants than in *daf-2* mutants (Fig. 6). *daf-2;daf-16* mutants have a decreased, normalized life span and a higher nutrient uptake and metabolism than the life-span-extended, hypo-metabolic *daf-2* mutants, which is also in agreement with the biological functions of the module genes.⁴¹

The top regulator of this module, CST-1, has a weight six times higher than the second regulator. Interestingly, its expression profiles are anticorrelated compared to the expression profiles of the genes in the module. CST-1 is a member of the STE20-like kinase family, which is known to mediate cell death triggered by oxidative stress in yeast and mammals.^{42,43}

In primary mammalian neurons, MST-1 mediates oxidative-stress-induced cell death by phosphorylating FOXO transcription factors, thereby disrupting their interaction with 14-3-3 proteins, promoting FOXO nuclear translocation and cell death.⁴³ Apparently, the interaction between STE20-like kinases and FOXO transcription factors upon oxidative stress is evolutionary conserved, since in *C. elegans*, CST-1 knock-down shortens life span and accelerates tissue aging through the forkhead transcription factor DAF-16.⁴³ Since mitochondria play an important role in the combat of oxidative stress and cell death,³⁷ both regulator and module genes seem to be involved in similar biological processes. Three genes in the module have been implicated in the determination of adult life span: while the knockdown of F26E4.6 and H28O16.1 both led to an extended life span, PRDX-2 knockdown reduced the life span of the organism. The genes in module 142 might also be targeted by DAF-16 and this is further supported by the different experiment partitions of experiment 20 *daf-2;daf-16* and *daf-2* mutants (Fig. 6). Except for *r1a-2*, whose expression pattern is also the least coherent of the module, all 5' intergenic sequences of the module genes contained the DAF-16 consensus binding element (data not shown), which has been shown to bind DAF-16 *in vitro* and to be present in many DAF-16 regulated genes.^{44,45} All 5' intergenic sequences included the DAF-16 associated element (data not shown), which is suggested to be an aging associated promoter element and to lead to DAF-16 dependent downregulation of transcription.^{41,45} The predicted transcription regulatory program of module 142 is that CST-1 activates DAF-16, which regulates the module genes (Fig. 7). DAF-16 is not predicted as a regulator by LeMoNe, because its expression profiles do not correspond to those of the module genes (data not shown), which can be explained by the fact that this transcription factor is broadly expressed and is involved in many biological processes (WormBase). Interestingly, CLR predicts DAF-3 as a regulator for four of the module genes and for three of the genes in module 137, a module that is also GO enriched for energy metabolism and growth. LeMoNe predicts three regulators for this module, but only the third regulator Y17G7B.20, which positively affects growth, has a

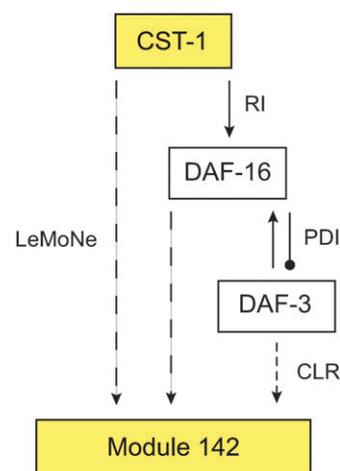


Fig. 7 The transcription regulatory network for module 142. The kinase CST-1 is predicted as top regulator. Experimental conditions in the condition clusters, external biological data and a reported regulatory interaction (RI) point to DAF-16 as intermediate transcription regulator. CLR predicted, reported regulatory (RI) and yeast one-hybrid (Y1H) interactions indicate that DAF-3 could also play a role in the path from CST-1 to the module genes.

clear biological relation with the module genes. DAF-3 positively regulates *daf-16* in the *C. elegans* TGF- β dauer pathway, which controls longevity through insulin signaling.⁴⁶ In addition, in yeast one-hybrid assays DAF-16 binds the *daf-3* promoter, hinting at a feedback control mechanism.¹⁷ This demonstrates again the complementariness of LeMoNe and CLR (Fig. 7). Module 142 may be part of the emerging picture that insulin-like molecules, through the activity of the DAF-2/insulin/IGF-I-like receptor, and the DAF-16/FKHRL1/FOXO transcription factor, control the ability of the organism to deal with oxidative stress, and interfere with metabolic programs that help to determine life span.⁴⁷

Conclusion

At a time when functional genomics data are increasing tremendously, reverse-engineering methods are gaining more and more importance. Such biological network inference methods are the interface between high- and low-throughput experiments, providing a means to bring structure and integration to data abundance, while putting forward selected follow-up experiments for experimental validation of hypotheses. In this study we have demonstrated that LeMoNe is a highly qualified reverse-engineering algorithm and capable of providing hypotheses on differential gene expression through inference of transcription regulatory networks from gene expression profiles.

Methods

Preprocessing of microarray data

Affymetrix expression profile data were assembled from ArrayExpress, GEO (gene expression omnibus) and through personal communications. We avoided the combination of different microarray platform data, since this could lead to

Table 5 Affymetrix microarray studies compiled in this study. Ref. = bibliographic reference

	Description	Database	Database ID	Ref.
1	Microarray-assisted positional cloning of TOM-1 and UNC-43	GEO	GSE2210	50
2	Lineage-specific regulatory network specified by PAL-1	GEO	GSE2180	51
3	The embryonic muscle transcriptome of <i>C. elegans</i>	GEO	GSE8462	52
4	Gene expression profiling of the <i>C. elegans</i> nervous system	GEO	GSE8004	53
5	A gene expression fingerprint of <i>C. elegans</i> embryonic motor neurons	GEO	GSE8159	54
6	Decline of nucleotide excision repair capacity in aged <i>C. elegans</i>	GEO	GSE4766	40
7	Developmental transcriptome profiling of the <i>C. elegans</i> pocket protein ortholog, <i>lin-37</i>	GEO	GSE6547	55
8	<i>C. elegans</i> gene expression in response to the pathogenic <i>P. aeruginosa</i> strain PA14	GEO	GSE5793	56
9	Genes regulated by PMK-1 and DAF-16 in a <i>daf-2(e1368)</i> background	GEO	GSE5801	56
10	Analysis of expression of genes regulated by DAF-19	GEO	GSE6563	57
11	Transcription profiling of the <i>C. elegans</i> RNAi defective mutants of <i>rde-1</i> and <i>rde-5</i>	ArrayExpress	E-MEXP-956	58
12	Transcription profiling of <i>C. elegans</i> <i>dcr-1</i> , <i>unc-32</i> homozygous mutants vs. coiling <i>unc-32</i> to investigate interference synthesis of small developmental RNAs	ArrayExpress	E-MEXP-957	58
13	Transcription profiling of <i>C. elegans</i> <i>sma-9</i> and <i>dbl-1</i> gene knockouts	ArrayExpress	E-MEXP-687	59
14	Transcription profiling of <i>C. elegans</i> after infection with <i>Microbacterium nematophilum</i>	ArrayExpress	E-MEXP-696	60
15	Expression profiling of single neuron types	From author	—	61
16	Translation of a small subset of <i>C. elegans</i> mRNAs is dependent on a specific eukaryotic translation initiation factor 4E isoform	From author	—	62
17	Interacting endogenous and exogenous RNAi pathways in <i>C. elegans</i>	From author	—	63
18	Monomethyl branched-chain fatty acids	From author	—	64
19	Unfolded protein response in <i>C. elegans</i>	From author	—	65
20	<i>C. elegans</i> dauer larvae and long-lived <i>daf-2</i> mutants	From author	—	41
21	Heme homeostasis is regulated by the conserved and concerted functions of HRG-1 proteins	GEO	GSE8696	66
22	Transcriptome profiling of <i>str-2</i> , <i>C. elegans</i> C2H2 Zn-finger	GEO	GSE9246	67
23	Expression data from wildtype and <i>gas-1</i> mitochondrial mutant <i>C. elegans</i>	GEO	GSE9896	68
24	Pairing competitive and topologically distinct regulatory modules enhances patterned gene expression	GEO	GSE9665	21
25	cRNA amplification methods enhance microarray identification of transcripts expressed in the nervous system	GEO	GSE9485	—
26	SKN-1-dependent oxidative stress response in <i>C. elegans</i>	GEO	GSE9301	—

erroneous conclusions. We gathered raw data (CEL files) for 26 experimental series, representing 489 arrays and 155 different conditions (Table 5). After preprocessing the microarray data in Bioconductor, R, with the robust multi-array average (RMA) method, a background-adjusted, normalized and summarized log-transformed expression value was obtained for each *C. elegans* probe set. An Affymetrix probe set is a set of 25-mer oligonucleotides (probes) whose sequence is designed to be complementary to the intended target gene. The probe set compositions in the Affymetrix chip description files (cdf) are considered static and do not take into account the continuously expanding knowledge about the transcriptome. A major concern is off-target hybridization, where probes target either a different gene than originally described, or target multiple genes.⁴⁸ Therefore, based on WS180, we created a custom *C. elegans* cdf file from the Affymetrix *C. elegans* cdf file that consisted of 16 723 probe sets of at least nine probes, each targeting with perfect sequence identity to its transcript and not aligning to any other gene's transcript with zero or one mismatch. In addition, we removed 5010 genes that showed no differential gene expression over all conditions (standard deviation lower than 0.5).

The public gene names in this paper are in accordance with WormBase version WS195.

LeMoNe analysis

We obtained a *C. elegans* gene expression profile compendium consisting of average expression values for 11 713 genes and 155 conditions. We used 1063 candidate regulators as input for LeMoNe (software available at

<http://bioinformatics.psb.ugent.be/software/details/LeMoNe>), mostly worm transcription factors (wTF2.1²⁶) (855), but also proteins with GO annotations “signal transduction”, “regulation of signal transduction” (187) and “chromatin modification” (6) and novel proteins that were found to bind DNA in yeast one-hybrid assays (15),^{15,17} and that were present on the original Affymetrix microarray compendium. We ran 20 independent Gibbs sampler LeMoNe runs, generating 20 local optima module cluster solutions, from which an ensemble-averaged solution of coexpression modules was created.⁸ We obtained 248 module clusters containing three or more genes. The program assigns each gene to only one cluster. Only C08E8.4 is present in two clusters, since compared to WormBase WS180, two genes merged into one in WS195. Next, LeMoNe predicted a ranked list of weighted regulators for each module, based on an ensemble of 10 regulatory program trees built using significant experiment sets found from 10 different experiment partitions and significant regulators sampled from 100 candidate regulator-split value pairs for each split between significant experiment clusters (see ref. 6 for more details on LeMoNe). The weight of a regulator is the sum of split scores over the different regulatory programs (10), for each regulator sampled (100) and for each level in the tree, taking into account the proportion of conditions covered. The split score (0–1) of a regulator indicates how well the expression-split value of the regulator explains the partition in conditions in the module. Theoretically, if for a specific module with for instance a regulatory tree of depth five, the same regulator has a perfect split score at all tree nodes and would be sampled at all instances, the maximum score would

be $1 \times 100 \times 5 \times 10 = 5000$. In general, due to the stochastic nature of the sampling and the imperfect split scores of most regulators, almost all tree nodes have multiple regulators assigned to them, especially at the lower levels in the regulatory tree, and the theoretical maximum score is never reached. We obtained a maximum score of 1224 for the topregulator SEX-1 in module 116 with a regulatory tree of five levels. We only considered regulators having a weight of 100 or higher (0.87% of all regulators assigned), resulting in on average four regulators per module and a total of 24012 predicted regulatory interactions. The weights of the regulators are comparable between the different modules.

Functional analysis on modules

The genes in each module were analyzed for GO enrichment with BiNGO,⁴⁹ using a custom annotation file with biological process annotations from Wormbase WS195 and a hypergeometric test, false discovery rate corrected for multiple testing with a confidence level of 95%. We kept the significant GO enrichment only if more than two genes in the module had the GO biological process annotation. In addition, the functional coherence of genes in the modules was independently assessed by identifying the percentage of genes in the module that shared functional gene–gene links in WormNet.²⁴ WormNet's coexpression data are all obtained from the Stanford Microarray Database, hence from cDNA microarrays.

CLR analysis

To the same microarray compendium we also applied CLR,² (software available at <http://gardnerlab.bu.edu/clr.html>). We retrieved mutual information z -scores for target gene interactions with all 1063 regulators. With a cut-off for the z -score at 5 (p -value $< 3 \times 10^{-7}$), we obtained a total of 14503 predicted regulatory interactions.

Reported regulatory interactions

We compared the predicted transcription regulatory interactions with 10734 non-autoregulatory known regulatory interactions in *C. elegans*, consisting of 598 yeast one-hybrid protein–DNA interactions,^{15,17,26} 525 Wormbase WS195 regulatory interactions (mostly regulator knockdown/knockout and target gene expression monitoring) and 9614 derived regulatory interactions from these two sets. We derived a regulatory interaction between protein A and target gene B if there was a path present through a combination of reported protein–DNA interactions and regulatory interactions from A to B, e.g. A binds C, C regulates D, D binds E and E regulates B. So the derived regulatory interactions are the “hidden” interactions present in the known yeast one-hybrid and regulatory interaction datasets.

Acknowledgements

We thank Kenny Billiau for establishing the web interface. Vanessa Vermeirssen is funded by a Belgian Federal Science Policy (BELSPO) Return Grant. Anagha Joshi is funded by an Early-Stage Marie Curie Fellowship. This work is supported by an IWT grant for the SBO Bioframe project and an IUAP grant for the BioMAGNet project (IAP P6/25).

References

- 1 M. Bansal, V. Belcastro, A. Ambesi-Impiombato and D. di Bernardo, *Mol. Syst. Biol.*, 2007, **3**, 78.
- 2 J. J. Faith, B. Hayete, J. T. Thaden, I. Mogno, J. Wierzbowski, G. Cottarel, S. Kasif, J. J. Collins and T. S. Gardner, *PLoS Biol.*, 2007, **5**, e8.
- 3 E. Segal, M. Shapira, A. Regev, D. Pe'er, D. Botstein, D. Koller and N. Friedman, *Nat. Genet.*, 2003, **34**, 166–176.
- 4 G. A. Wray, M. W. Hahn, E. Abouheif, J. P. Balhoff, M. Pizer, M. V. Rockman and L. A. Romano, *Mol. Biol. Evol.*, 2003, **20**, 1377–1419.
- 5 A. J. Whitmarsh and R. J. Davis, *Cell Mol. Life Sci.*, 2000, **57**, 1172–1183.
- 6 A. Joshi, R. De Smet, K. Marchal, Y. Van de Peer and T. Michoel, *Bioinformatics*, 2009, **25**, 490–496.
- 7 T. Michoel, S. Maere, E. Bonnet, A. Joshi, Y. Saey, T. Van den Bulcke, K. Van Leemput, P. van Remortel, M. Kuiper, K. Marchal and Y. Van de Peer, *BMC Bioinf.*, 2007, **8**(suppl. 2), S5.
- 8 A. Joshi, Y. Van de Peer and T. Michoel, *Bioinformatics*, 2008, **24**, 176–183.
- 9 T. Michoel, R. De Smet, A. Joshi, Y. Van de Peer and K. Marchal, *BMC Syst. Biol.*, 2009, **3**, 49.
- 10 I. Antoshechkin and P. W. Sternberg, *Nat. Rev. Genet.*, 2007, **8**, 518–532.
- 11 S. Balaji, M. M. Babu, L. M. Iyer, N. M. Luscombe and L. Aravind, *J. Mol. Biol.*, 2006, **360**, 213–227.
- 12 S. Gama-Castro, V. Jimenez-Jacinto, M. Peralta-Gil, A. Santos-Zavaleta, M. I. Penaloza-Spinola, B. Contreras-Moreira, J. Segura-Salazar, L. Muniz-Rascado, I. Martinez-Flores, H. Salgado, C. Bonavides-Martinez, C. Abreu-Goodger, C. Rodriguez-Penagos, J. Miranda-Rios, E. Morett, E. Merino, A. M. Huerta, L. Trevino-Quintanilla and J. Collado-Vides, *Nucleic Acids Res.*, 2008, **36**, D120–D124.
- 13 H. Lei, J. Liu, T. Fukushige, A. Fire and M. Krause, *Development (Cambridge, U. K.)*, 2009, **136**, 1241–1249.
- 14 S. W. Oh, A. Mukhopadhyay, B. L. Dixit, T. Raha, M. R. Green and H. A. Tissenbaum, *Nat. Genet.*, 2006, **38**, 251–257.
- 15 B. Deplancke, A. Mukhopadhyay, W. Ao, A. M. Elewa, C. A. Grove, N. J. Martinez, R. Sequerra, L. Doucette-Stamm, J. S. Reece-Hoyes, I. A. Hope, H. A. Tissenbaum, S. E. Mango and A. J. Walhout, *Cell*, 2006, **125**, 1193–1205.
- 16 N. J. Martinez, M. C. Ow, M. I. Barrasa, M. Hammell, R. Sequerra, L. Doucette-Stamm, F. P. Roth, V. R. Ambros and A. J. Walhout, *Genes Dev.*, 2008, **22**, 2535–2549.
- 17 V. Vermeirssen, M. I. Barrasa, C. A. Hidalgo, J. A. Babon, R. Sequerra, L. Doucette-Stamm, A. L. Barabasi and A. J. Walhout, *Genome Res.*, 2007, **17**, 1061–1071.
- 18 O. Hobert, *Cold Spring Harbor Symp. Quant. Biol.*, 2006, **71**, 181–188.
- 19 T. Inoue, M. Wang, T. O. Ririe, J. S. Fernandes and P. W. Sternberg, *Proc. Natl. Acad. Sci. U. S. A.*, 2005, **102**(14), 4972–4977.
- 20 T. O. Ririe, J. S. Fernandes and P. W. Sternberg, *Proc. Natl. Acad. Sci. U. S. A.*, 2008, **105**, 20095–20099.
- 21 I. Yanai, L. R. Baugh, J. J. Smith, C. Roehrig, S. S. Shen-Orr, J. M. Claggett, A. A. Hill, D. K. Slonim and C. P. Hunter, *Mol. Syst. Biol.*, 2008, **4**, 163.
- 22 Y. V. Budovskaya, K. Wu, L. K. Southworth, M. Jiang, P. Tedesco, T. E. Johnson and S. K. Kim, *Cell*, 2008, **134**, 291–303.
- 23 R. X. Yu, J. Liu, N. True and W. Wang, *PLoS One*, 2008, **3**, e1821.
- 24 I. Lee, B. Lehner, C. Crombie, W. Wong, A. G. Fraser and E. M. Marcotte, *Nat. Genet.*, 2008, **40**, 181–188.
- 25 A. Rogers, I. Antoshechkin, T. Bieri, D. Blasiar, C. Bastiani, P. Canaran, J. Chan, W. J. Chen, P. Davis, J. Fernandes, T. J. Fiedler, M. Han, T. W. Harris, R. Kishore, R. Lee, S. McKay, H. M. Muller, C. Nakamura, P. Ozersky, A. Petcherski, G. Schindelman, E. M. Schwarz, W. Spooner, M. A. Tuli, K. Van Auken, D. Wang, X. Wang, G. Williams, K. Yook, R. Durbin, L. D. Stein, J. Spieth and P. W. Sternberg, *Nucleic Acids Res.*, 2008, **36**, D612–D617.
- 26 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 and A. J. Walhout, *Nat. Methods*, 2007, **4**, 659–664.

- 27 D. Greenstein, *WormBook*, 2005, 1–12.
- 28 T. C. Evans and C. P. Hunter, *WormBook*, 2005, 1–11.
- 29 P. Gonczy and L. S. Rose, *WormBook*, 2005, 1–20.
- 30 A. Antebi, *WormBook*, 2006, 1–13.
- 31 V. Reinke, I. S. Gil, S. Ward and K. Kazmer, *Development (Cambridge, U. K.)*, 2004, **131**, 311–323.
- 32 A. L. Lublin and T. C. Evans, *Dev. Biol.*, 2007, **303**, 635–649.
- 33 M. Shimada, H. Yokosawa and H. Kawahara, *Genes Cells*, 2006, **11**, 383–396.
- 34 R. Lin, *Dev. Biol.*, 2003, **258**, 226–239.
- 35 M. R. Detwiler, M. Reuben, X. Li, E. Rogers and R. Lin, *Dev. Cell*, 2001, **1**, 187–199.
- 36 C. M. Phillips and A. F. Dernburg, *Dev. Cell*, 2006, **11**, 817–829.
- 37 P. Kakkar and B. K. Singh, *Mol. Cell. Biochem.*, 2007, **305**, 235–253.
- 38 M. Olahova, S. R. Taylor, S. Khazaipoul, J. Wang, B. A. Morgan, K. Matsumoto, T. K. Blackwell and E. A. Veal, *Proc. Natl. Acad. Sci. U. S. A.*, 2008, **105**, 19839–19844.
- 39 T. R. Golden, K. B. Beckman, A. H. Lee, N. Dudek, A. Hubbard, E. Samper and S. Melov, *Aging Cell*, 2007, **6**, 179–188.
- 40 J. N. Meyer, W. A. Boyd, G. A. Azzam, A. C. Haugen, J. H. Freedman and B. Van Houten, *GenomeBiology*, 2007, **8**, R70.
- 41 J. J. McElwee, E. Schuster, E. Blanc, J. H. Thomas and D. Gems, *J. Biol. Chem.*, 2004, **279**, 44533–44543.
- 42 S. H. Ahn, W. L. Cheung, J. Y. Hsu, R. L. Diaz, M. M. Smith and C. D. Allis, *Cell*, 2005, **120**, 25–36.
- 43 M. K. Lehtinen, Z. Yuan, P. R. Boag, Y. Yang, J. Villen, E. B. Becker, S. DiBacco, N. de la Iglesia, S. Gygi, T. K. Blackwell and A. Bonni, *Cell*, 2006, **125**, 987–1001.
- 44 T. Furuyama, T. Nakazawa, I. Nakano and N. Mori, *Biochem. J.*, 2000, **349**, 629–634.
- 45 C. T. Murphy, S. A. McCarroll, C. I. Bargmann, A. Fraser, R. S. Kamath, J. Ahringer, H. Li and C. Kenyon, *Nature*, 2003, **424**, 277–283.
- 46 W. M. Shaw, S. Luo, J. Landis, J. Ashraf and C. T. Murphy, *Curr. Biol.*, 2007, **17**, 1635–1645.
- 47 R. Baumeister, E. Schaffitzel and M. Hertweck, *J. Endocrinol.*, 2006, **190**, 191–202.
- 48 T. Casneuf, Y. Van de Peer and W. Huber, *BMC Bioinf.*, 2007, **8**, 461.
- 49 S. Maere, K. Heymans and M. Kuiper, *Bioinformatics*, 2005, **21**, 3448–3449.
- 50 M. Dybbs, J. Ngai and J. M. Kaplan, *PLoS Genet.*, 2005, **1**, 6–16.
- 51 L. R. Baugh, A. A. Hill, J. M. Claggett, K. Hill-Harfe, J. C. Wen, D. K. Slonim, E. L. Brown and C. P. Hunter, *Development (Cambridge, U. K.)*, 2005, **132**, 1843–1854.
- 52 R. M. Fox, J. D. Watson, S. E. Von Stetina, J. McDermott, T. M. Brodigan, T. Fukushige, M. Krause and D. M. Miller 3rd., *GenomeBiology*, 2007, **8**, R188.
- 53 S. E. Von Stetina, J. D. Watson, R. M. Fox, K. L. Olszewski, W. C. Spencer, P. J. Roy and D. M. Miller 3rd., *GenomeBiology*, 2007, **8**, R135.
- 54 R. M. Fox, S. E. Von Stetina, S. J. Barlow, C. Shaffer, K. L. Olszewski, J. H. Moore, D. Dupuy, M. Vidal and D. M. Miller 3rd., *BMC Genomics*, 2005, **6**, 42.
- 55 N. V. Kirienko and D. S. Fay, *Dev. Biol.*, 2007, **305**, 674–684.
- 56 E. R. Troemel, S. W. Chu, V. Reinke, S. S. Lee, F. M. Ausubel and D. H. Kim, *PLoS Genet.*, 2006, **2**, e183.
- 57 N. Chen, A. Mah, O. E. Blacque, J. Chu, K. Phgora, M. W. Bakhoum, C. R. Newbury, J. Khattra, S. Chan, A. Go, E. Efimenko, R. Johnsen, P. Phirke, P. Swoboda, M. Marra, D. G. Moerman, M. R. Leroux, D. L. Baillie and L. D. Stein, *GenomeBiology*, 2006, **7**, R126.
- 58 N. C. Welker, J. W. Habig and B. L. Bass, *RNA*, 2007, **13**, 1090–1102.
- 59 J. Liang, L. Yu, J. Yin and C. Savage-Dunn, *Dev. Biol.*, 2007, **305**, 714–725.
- 60 D. O'Rourke, D. Baban, M. Demidova, R. Mott and J. Hodgkin, *Genome Res.*, 2006, **16**, 1005–1016.
- 61 M. E. Colosimo, A. Brown, S. Mukhopadhyay, C. Gabel, A. E. Lanjuin, A. D. Samuel and P. Sengupta, *Curr. Biol.*, 2004, **14**, 2245–2251.
- 62 T. D. Dinkova, B. D. Keiper, N. L. Korneeva, E. J. Aamodt and R. E. Rhoads, *Mol. Cell. Biol.*, 2005, **25**, 100–113.
- 63 R. C. Lee, C. M. Hammell and V. Ambros, *RNA*, 2006, **12**, 589–597.
- 64 M. Kniazeva, Q. T. Crawford, M. Seiber, C. Y. Wang and M. Han, *PLoS Biol.*, 2004, **2**, e257.
- 65 X. Shen, R. E. Ellis, K. Sakaki and R. J. Kaufman, *PLoS Genet.*, 2005, **1**, e37.
- 66 A. Rajagopal, A. U. Rao, J. Amigo, M. Tian, S. K. Upadhyay, C. Hall, S. Uhm, M. K. Mathew, M. D. Fleming, B. H. Paw, M. Krause and I. Hamza, *Nature*, 2008, **453**, 1127–1131.
- 67 N. V. Kirienko, J. D. McEnerney and D. S. Fay, *PLoS Genet.*, 2008, **4**, e1000059.
- 68 M. J. Falk, Z. Zhang, J. R. Rosenjack, I. Nissim, E. Daikhin, I. Nissim, M. M. Sedensky, M. Yudkoff and P. G. Morgan, *Mol. Genet. Metab.*, 2008, **93**, 388–397.