

Executing multicellular differentiation: quantitative predictive modelling of *C.elegans* vulval development

Nicola Bonzanni^{1,2}, Elzbieta Krepska², K. Anton Feenstra^{1,*}, Wan Fokkink², Thilo Kielmann², Henri Bal² and Jaap Heringa¹

¹Centre for Integrative Bioinformatics and ²Department of Computer Science, Vrije Universiteit, De Boelelaan 1083, 1081 HV Amsterdam, The Netherlands

Received on October 22, 2008; revised on May 25, 2009; accepted on June 4, 2009

Advance Access publication June 10, 2009

Associate Editor: Olga Troyanskaya

ABSTRACT

Motivation: Understanding the processes involved in multi-cellular pattern formation is a central problem of developmental biology, hopefully leading to many new insights, e.g. in the treatment of various diseases. Defining suitable computational techniques for development modelling, able to perform *in silico* simulation experiments, is an open and challenging problem.

Results: Previously, we proposed a coarse-grained, quantitative approach based on the basic Petri net formalism, to mimic the behaviour of the biological processes during multicellular differentiation. Here, we apply our modelling approach to the well-studied process of *Caenorhabditis elegans* vulval development. We show that our model correctly reproduces a large set of *in vivo* experiments with statistical accuracy. It also generates gene expression time series in accordance with recent biological evidence. Finally, we modelled the role of microRNA mir-61 during vulval development and predict its contribution in stabilizing cell pattern formation.

Contact: feenstra@few.vu.nl

Supplementary information: Supplementary data are available at *Bioinformatics* online.

1 INTRODUCTION

Many efforts have been undertaken to elucidate how cells are able to coordinate different and sometimes conflicting signals, producing a precise phenotype during the animal organogenesis (Sternberg, 2005). *Caenorhabditis elegans* vulval development provides an elegant and relatively well-charted model to study how multiple pathways, in multiple cells, interact to produce developmental patterns.

The *C.elegans* hermaphrodite vulva develops from three of the six vulval precursor cells (VPCs), consecutively numbered from P3.p to P8.p in Figure 1. Each VPC is competent to respond to intercellular signals, and is potentially able to adopt either of the three cell fates: 1°, 2° or 3°. Each fate corresponds to a specific cell division pattern. The 1° and 2° fate cell lineages constitute the vulva, generating eight and seven progeny cells, respectively. The 3° fate lineage becomes a constituent of the hyp7 hypodermal syncytium, a large cell-like structure with many nuclei enveloping the developing nematode.

*To whom correspondence should be addressed.

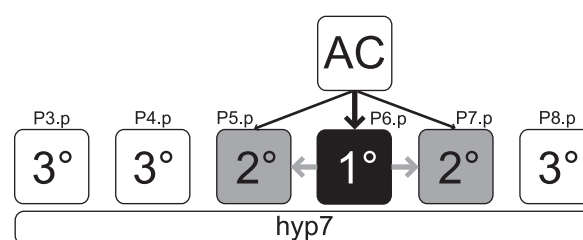


Fig. 1. Vulval development in the wild-type *C.elegans*, showing the AC, the VPCs (P3.p–P8.p) and the hyp7. The inductive signal from the AC promotes the 1° fate in P6.p, and stimulates the production of the lateral signal near the flanking cells, promoting the 2° fate in P5.p and P7.p. The 3° fate lineage becomes a constituent of the hyp7.

In the wild-type hermaphrodite, the six VPCs adopt an invariant 3°–3°–2°–1°–2°–3° pattern (Sternberg and Horvitz, 1986), shown in Figure 1. This precise fate distribution is the result of the interplay between two competing signals: the spatially graded inductive signal produced by the anchor cell (AC), and the lateral signal originating from a presumptive 1° fate cell.

During this cell–cell interaction, the inductive epidermal growth-factor signal is produced by the AC and transported to the three nearest precursor cells. The signal is encoded by the protein LIN-3 and transduced by the receptor LET-23 into the Ras/MAPK pathway. This has the direct effect of up-regulating MPK-1, and of promoting 1° fate in P6.p. Further downstream the Ras/MAPK pathway, LIN-12 is down-regulated (Shaye and Greenwald, 2002) to suppress the promotion of 2° fate, while production of the lateral signal is stimulated (Chen and Greenwald, 2004). This signal promotes 2° fate in the neighbouring cells P5.p and P7.p (Sundaram, 2004), and inhibits Ras signalling to block transduction of the inductive signal through the Ras/MAPK pathway (Yoo *et al.*, 2004). This negative feedback helps to maximize LIN-12 activity in the presumptive 2° fate cells (Yoo and Greenwald, 2005).

The first diagrammatic model, describing the regulatory network underlying VPC determination, was proposed by Sternberg and Horvitz (1989). Since then, global understanding of the biological network has improved greatly. The first computational model, proposed by Kam *et al.* (2003), combined multiple experimental ‘scenarios’ from Sternberg and Horvitz (1986) into a single model, using Live Sequence Charts (LSCs). Afterwards, in two landmark

papers, Fisher *et al.* (2005, 2007) suggested two state-based mechanistic models. The first (Fisher *et al.*, 2005) used statecharts to represent internal states of components, and LSCs to execute actions between them. They formalized Sternberg's model (Sternberg and Horvitz, 1989), but did not incorporate any additional data. A more recent approach (Fisher *et al.*, 2007) was based on Reactive Modules, with modelling principles akin to the previous paper. In contrast to the model presented in the current article, the three listed models build on representing rules that the system adheres to, rather than modelling the underlying biological processes. Two other insightful models of *C.elegans* vulval development have been published. Giurumescu *et al.* (2006) proposed a partial model based on ODEs, while Sun and Hong (2007) developed a model based on automatically learned dynamic Bayesian networks with discrete states. Independent from us, Li *et al.* (2009) recently modelled part of *C.elegans* vulval development using hybrid functional Petri nets with extensions. While they focused on model validation, we additionally generated new insightful predictions.

In this article, we apply our approach (Krepska *et al.*, 2008), which is discrete, non-deterministic and based on Petri nets to *C.elegans* vulval development. Petri nets are a convenient formalism to represent biological networks. This formalism models process synchronization, asynchronous events, conflicts and in general concurrent systems in a natural way. Moreover, Petri nets offer direct insights into causal relationships, and allow a graphical visualization that resembles the diagrams used to describe biological knowledge. The reader may find recent survey papers concerning modelling of biological systems with Petri nets in Koch and Heiner (2008), Chaouiya (2007), Matsuno *et al.* (2006) and Peleg *et al.* (2005).

Several adaptations of the Petri net formalism have been introduced in the context of modelling biological systems. On the one hand, qualitative Petri nets (Gilbert *et al.*, 2007) can be used for structural and invariant analysis, but they greatly abstract from the biological system. On the other hand, stochastic Petri nets (Goss and Peccoud, 1998) incorporate kinetic constants, but these are mostly unknown or approximate. Hybrid Petri nets (Matsuno *et al.*, 2000) and their extensions on which Cell Illustrator (Matsuno *et al.*, 2006) is based, are rich and expressive, but model understanding and causal backtracking are impeded by the complexity of the formalism.

In our model, we have chosen to preserve the simplicity of the original Petri net formalism. Our modelling approach is aimed to mimic the underlying biological mechanisms as much as possible, and not only to reproduce the expected phenotype according to a specific set of mutations. To achieve this, we apply a principle of maximal parallelism (Burkhard, 1980), and bounded execution with overshooting (Krepska *et al.*, 2008). Using this simple framework, we identified different modules, each corresponding to different biological functions. Thus, combining functional modules into cells, and joining such cells together, we iteratively developed the whole network. Unlike the aforementioned works on formal modelling of *C.elegans* vulval development, the ability of our model to capture biological functions into small building blocks allows these to be reused in new case studies on multi-cellular signalling and regulation modelling.

We show that our model, encoding biological hypotheses from the literature (Shaye and Greenwald, 2002; Yoo and Greenwald, 2005), is able to reproduce *in silico* a set of *in vivo* experiments, providing the necessary statistical data to establish a more detailed comparison with biological observations than was previously possible. To the

best of our knowledge, we are the first to model microRNA interactions during *C.elegans* vulval development. Furthermore, we predict a possible 'tuning' role played by the *mir-61* microRNA gene, in ensuring stability of the fate pattern.

2 METHODS

2.1 Biological interpretation of quantitative Petri nets

A Petri net (Petri, 1962; Reisig and Rozenberg, 1998) is a bipartite-directed graph consisting of two kinds of nodes: places that indicate the local availability of resources, and transitions which are active components that can change the state of the resources. Each place can hold one or more tokens. Weighted arcs connect places and transitions.

Instead of further enriching the formalism to extend its expressiveness (but also its complexity), we focus on preserving the simplicity of the formalism, and develop an execution semantics which resembles biology. In Krepska *et al.* (2008), we explain a method to represent biological knowledge as a Petri net. Places represent genes, protein species and complexes, while transitions represent biological processes. Firing of a transition is execution of a process, e.g. consuming substrates or creating products.

The number of tokens in our model does not represent directly the number of molecules of proteins or a fixed molar concentration as in Gilbert *et al.* (2007). In our model, we interpret this number in two ways. For genes as a Boolean value, 0 means not present and 1 present. For proteins, we use abstract concentration levels 0–6: going from not present, via low, medium and high concentration to saturated level. The rationale behind this approach is to abstract away from unknown absolute molecule concentration levels, as we intend to represent relative concentrations. We choose to use seven concentration levels in order to stay in between a simple Boolean level and a complex ODE model, and because seven concentration levels sufficed to express the biological knowledge from the literature on *C.elegans* vulva development in a satisfactory fashion. If desired, a modeller could fine-tune the granularity of the model by adjusting the number of available concentration levels.

Biological systems are highly concurrent, as in cells all reactions can happen in parallel and most are independent of each other. Therefore, we apply a principle of maximal parallelism (Burkhard, 1980). A fully asynchronous approach would allow one part of the network to deploy prolonged activity, while another part of the network shows no activity at all. In real life, all parts can progress at roughly the same 'speed' (Fisher *et al.*, 2008). Maximal parallelism promotes activity throughout the network, so that values on arcs really capture relative speed and concentration levels, as corroborated by our experiments. The maximal parallel execution semantics can be summarized informally as *execute greedily as many transitions as possible in one step*. A step S is a multi-set of transitions, i.e. a transition can occur multiple times in S . A maximally parallel step is a step that leaves no enabled transitions in the net, and in principle should be developed in such a way that it corresponds to one time step in the evolution of the biological system. This is possible because the modeller can capture relative speeds using appropriate weights on arcs. Typically, if in one time unit a protein A is produced four times more than a protein B, then the transition that captures production of A should have a weight that is four times as large as the weight of the one that captures B production. In case more than one maximally parallel step is possible, one is selected randomly. In short, implementing a pure maximally parallel semantics requires to generate all possible partitions of tokens, and select one randomly, uniformly. However, with the growth of the network, this procedure becomes prohibitively slow. Therefore, we approximated it by building a maximally parallel step incrementally, selecting one transition after another, randomly, until all enabled transitions have been exhausted, as explained in Krepska *et al.* (2008).

Unrestricted production of proteins is usually not realistic, as in nature the cell would saturate with the product, and the reaction would slow down or stop. Therefore, to mimic this behaviour, each place has a predefined

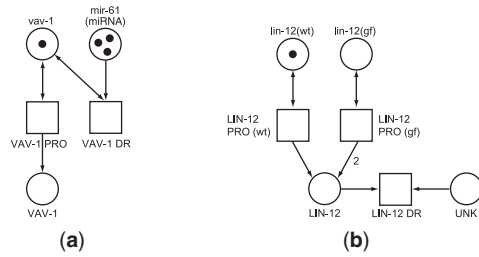


Fig. 2. (a) The presence of the microRNA mir-61 down-regulates VAV-1, by enabling the transition VAV-1 DR, which is in conflict with VAV-1 PRO. (b) Two example modules, gene production (vertical) and endocytosis (horizontal), interact in the Notch/LIN-12 pathway.

maximum capacity $\mathcal{N}=6$. To guarantee that the highest concentration level can be attained, we introduced bounded execution with overshooting. A transition can only fire if each output place holds fewer than \mathcal{N} tokens. Since each transition can possibly move more than one token at once into its output places, each transition can overshoot the pre-given capacity \mathcal{N} at most once. Therefore, the network is bounded with a finite bound $k \geq \mathcal{N}$.

2.2 Model construction

We developed an executable Petri net model for cell fate determination during *C.elegans* vulval induction. This large network can be visualized on the web page of our project (<http://www.cs.vu.nl/concell>); a schematic representation is given in Figure 5. The entire network comprises 600 nodes (places and transitions) and 1000 arcs. Nevertheless the simplicity of the formalism, and its graphical representation, helps us to identify different modules. These correspond to different biological functions, such as gene expression, protein activation and protein degradation. It is possible to reuse modules corresponding to a function, like small building blocks, to compose more complex modules, and eventually build a full cell. The cell itself is a module that can be reused. Applying these principles, we have built the VPC network out of six interconnected cells as identical modules of a multi-potent cell. We also built a separate block for the AC (producing the inductive signal) and for the hyp7. The possibility to divide the entire graph into simple, small and meaningful modules has three main advantages: (i) the modelling process becomes easier, (ii) the resulting network is homogeneous, and (iii) modules (at different levels) can be reused throughout the model, or for modelling other organisms.

Figure 2 shows selected examples of how to represent biological modules as a Petri net. Figure 2a illustrates VAV-1 down-regulation by decreasing the translation rate of the gene *vav-1*. In fact, if mir-61 is not present, the reaction VAV-1 PRO is enabled and produces the protein. However, when mir-61 is present, the reaction VAV-1 DR is enabled and has 0.5 chance of firing compared with VAV-1 PRO, thus the production of VAV-1 will halve. Figure 2b depicts two connected basic modules, a gene expression and the endocytosis mediated down-regulation of LIN-12. In this example, activation of the Ras/MAPK cascade leads to the transcription of a hitherto unknown gene that enhances the LIN-12 endocytosis, as hypothesized by Shaye and Greenwald (2002). Note that here the produced LIN-12 is removed, while in Figure 2a, the gene production was reduced. An alternative way to represent down-regulation using transitions has been proposed by Grunwald *et al.* (2008).

At the system level, a module can be viewed as a ‘meta-transition’, a Petri net with specified inputs (places which can receive tokens) and outputs (arcs outgoing of the module). Instead of constructing the Petri net model using Boolean functions as basic network components (Sackmann *et al.*, 2006) and subsequently check the biological meaning of each subnetwork extracted by invariant analysis (Grunwald *et al.*, 2008; Sackmann *et al.*, 2006), we focused on using basic biological functions as network building

Table 1. Description of modules constituting the model of *C.elegans* vulval development depicted in Figures 4 and 5

Module	Function
SEM-5	Production and activation of SEM-5 from gene <i>sem-5</i> (wt).
LET-60	Production and activation of LET-60 from gene <i>let-60</i> (wt).
LIN-3	Reception of LIN-3 from AC and hyp7.
LET-23	Production and activation of LET-23 from gene <i>let-23</i> (wt). Down-regulation of LET-23 by DPY-23.
LST	Production and activation of LSTs from <i>lst-1</i> (wt), <i>lst-2</i> (wt) and <i>lst-4</i> (wt) genes. Down-regulation of LSTs by MPK-1. Up-regulation of LSTs promoted by LIN-12*.
MPK-1	Production and activation of MPK-1 from <i>mpk-1</i> (wt) gene. Down-regulation of MPK-1 by LST.
DSL	Production of DSL signal.
LATERAL	Transport of lateral signal (DSL) to adjacent cells.
DPY-23	Production of DPY-23 from <i>dpy-23</i> (wt) gene, promoted by LIN-12*.
LIN-12	Production of LIN-12 from <i>lin-12</i> (wt) gene. Activation of LIN-12 by binding to DSL. Endocytotic down-regulation of LIN-12 mediated by VAV-1 and promoted by the Ras/MAPK pathway.
VAV-1	Production of VAV-1 from <i>vav-1</i> (wt) gene. Down-regulation of VAV-1 by microRNA mir-61.
MIR-61	Production of miR-61 microRNA.
AC	Production of LIN-3 and diffusion in a graded fashion to P6.p and the two adjacent cells P5.p and P7.p.
hyp7	Production of LIN-3 and diffusion to all VPCs.
Not shown	Constitutive degradation of various proteins.

Protein names followed by a ‘*’ stand for the active proteins.

blocks. In our experience, the procedure of building a modular biological Petri net can be split into five phases:

Level 1: basic biological functions. We created six basic modules representing the basic biological functions used to encode the relations described in the literature related to *C.elegans* vulva development: protein production, protein activation, down-regulation, up-regulation, signalling and constitutive degradation.

Level 2: protein interactions. Combining basic modules, we built more complex blocks, each modelling the interactions of one protein. The division into protein interaction modules is presented in Table 1. Figure 3 shows an example of how basic biological functions are combined to build protein interaction modules.

Level 3: pathways. In Figure 4, modules LIN-3, LET-23, SEM-5, LET-60, MPK-1 and DSL constitute the Ras/MAPK pathway, and modules LIN-12, VAV-1, MIR-61, DPY-23 and LST constitute the competing Notch/LIN-12 pathway.

Level 4: cells. Figure 4 presents the Petri net model of a single VPC cell with four links to the environment.

Level 5: multi-cellular interactions. In Figure 5, we show how the six VPCs, AC and hyp7 modules are connected. Adjacent cells are linked with each other, the hyp7 connects to all six cells, and the AC can directly influence cells P5.p, P6.p and P7.p.

Figure 3 highlights the top-left portion of the VPC model depicted in Figure 4. One can see how basic biological functions are reused in different protein interaction modules, where the links describe the interactions between different modules. For instance, in Figure 3, the LET-23 module is connected to LIN-3, which is connected to SEM-5, which in turn interacts with LET-60. The biological mechanisms underlying these interactions

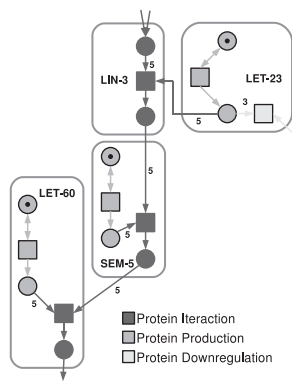


Fig. 3. Basic biological functions used in more complex modules.

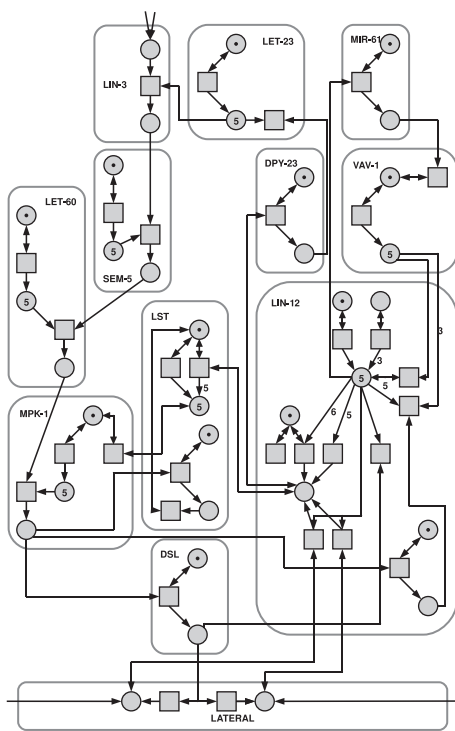


Fig. 4. Schematic representation of a VPC in our Petri net model. Each rounded box is a module. Note that LIN-3 and LATERAL are connected to the environment.

are found in the literature and encoded by the basic biological functions mentioned. The network shown in Figure 3 models the first steps during signal transduction within the Ras/MAPK cascade, where the transmembrane receptor LET-23 is activated by the ligand LIN-3. The resulting activated complex then activates the core Ras protein LET-60, by signalling through SEM-5.

2.3 Modelling genetic perturbations

For each genetic background, each gene can be in wild-type form (*wt*, i.e. the most common form of a gene as it occurs in nature) or in one of the following mutated forms: loss-of-function (*lf*, e.g. the gene is deleted or dysfunctional) or gain-of-function (*gf*, e.g. the gene transcription is over-stimulated). It is possible to derive an initial configuration corresponding

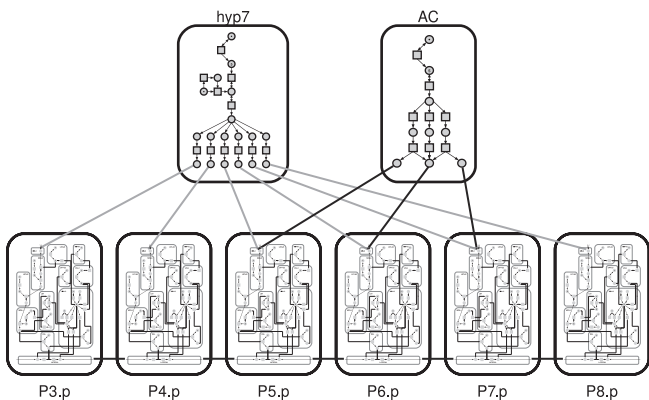


Fig. 5. Schematic representation of the whole system. The VPCs are connected with the AC, the hyp7 and their adjacent cells.

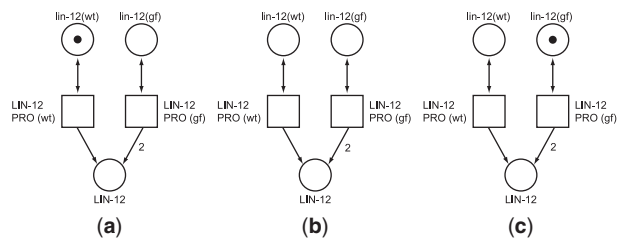


Fig. 6. Gene expression with different initial conditions, corresponding to different genetic backgrounds. (a) Corresponds to *lin12* wild-type, (b) corresponds to the *lin12(lf)* mutant and (c) corresponds to the *lin12(gf)* mutant. Note that arcs with two heads represent two arcs, one either way.

to a given genetic perturbation placing a token in one of the two different places used to represent gain-of-function and wild-type for each gene in the genetic background. Loss-of-function mutation is represented by token removal. It is therefore possible to initiate the network in an appropriate initial configuration by simply placing or removing tokens in opportune places.

Figure 6a depicts an example of a typical gene transcription. The transition LIN-12 PRO(*wt*) produces LIN-12 proteins when the wild-type gene *lin-12(wt)* is present. When the gene is not present [i.e. *lin-12(wt)* holds no token], the event does not take place. Figure 6b and c depicts two different genetic backgrounds, corresponding respectively, to the loss-of-function and gain-of-function of the *lin-12* gene.

2.4 Model calibration

We started off with assuming that initially proteins are expressed at low basal levels and reactions require high protein concentration levels. Therefore, we set the initial concentration levels for proteins to zero and we assigned high requirements and low-level production to all transitions, respectively, arc weights five and one (see the Supplementary Material for an example).

We subsequently simulated the 22 *in vivo* experiments in our calibration set (Table 2). We identified mismatches between the simulation results and the expected phenotypes, and back-tracked the problem (e.g. an overly strong or weak down-regulation) following the causal chain from one module to the other (i.e. from products, to transitions, up to their requirements). For selected modules, arc weights and occasionally initial protein concentration levels were fine-tuned to recover the expected behaviour.

This manual calibration process iteratively converged upon a stable and fixed set of parameters that we used for all further simulations. During the process, we noticed that only in very few cases single parameter adjustments were able to sensibly change the simulation results, whereas

Table 2. *In vivo* experiments selected in Fisher *et al.* (2005), and used by us for model calibration

Exp.	AC Genotype		Fate Pattern								Ref.
	lst	Vul	lin-15	lin-12	P3.p	P4.p	P5.p	P6.p	P7.p	P8.p	
1	+				3	3	2	1	2	3	a
2	+	lf			3	3	1	1	1	3	b
3	+		lf		3	3	3	3	3	3	c
5	+			lf	1\2	1\2	2	1	2	1\2	c
6	+	lf		lf	1	1	1	1	1	1	d
7	+		lf	lf	3	3	3	3	3	3	e
9	+			lf	3	3	1	1	1	3	f
10	+	lf		lf	3	3	1	1	1	3	g
11	+		lf		lf	3	3	3	3	3	c
13	+			lf	1	1	1	1	1	1	c
17	+			gf	2	2	2	1	2	2	c
19	+		lf	gf	2	2	2	2	2	2	c
21	+			lf	1\2	1\2	2	1	2	1\2	c
25	—				3	3	3	3	3	3	h
26	—	lf			3	3	3	3	3	3	d
29	—			lf	1\2	1\2	1\2	1\2	1\2	1\2	c
33	—			lf	3	3	3	3	3	3	c
37	—			lf	1	1	1	1	1	1	c
41	—			gf	2	2	2	2	2	2	c
42	—	lf		gf	2	2	2	2	2	2	d
43	—		lf	gf	2	2	2	2	2	2	c
45	—			lf	1\2	1\2	1\2	1\2	1\2	1\2	c

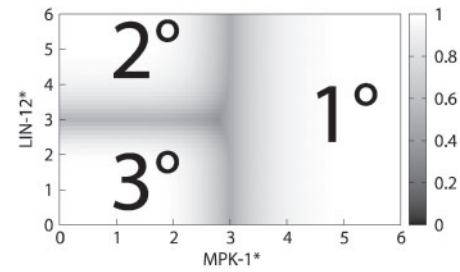
a: Sulston and Horvitz (1977); b: Berset *et al.* (2005); Yoo *et al.* (2004); c: Sternberg and Horvitz (1989); d: Berset *et al.* (2001); e: Cui *et al.* (1977); Ferguson and Horvitz (1989); Sternberg and Horvitz (1989); f: Greenwald *et al.* (1983); Sternberg and Horvitz (1989); g: Berset and Hajnal (unpublished data); h: Kimble (1981). In the AC column, '—' stands for no AC, while '+' means that the AC is present. In the Genotype column, for each gene a loss-of-function (*lf* or knock-out) or gain-of-function (*gf* or overexpression) mutation is indicated. *lst* is the group of *lst-1*, *lst-2*, *lst-3*, *lst-4* and *dpy-23*. *Vul* is the group of *let-23*, *sem-5*, *let-60* and *mpk-1*. In the Fate Pattern column, 1 indicates 1° cell fate, 2, 2° fate, 3, 3° fate, and 1\2 either 1° or 2° fate.

more often combinations of parameters were changed to approach the expected behaviour. This suggests a 'spectrum of sensitivities' as discussed in Gutenkunst *et al.* (2007) that should allow the modellers to focus on predictions rather than on parameters.

2.5 Simulation procedure

In experimental biology, experimental replicates are necessary to overcome the variability intrinsic to biological systems. In our modelling approach, which is non-deterministic, we interpret the outcome of a simulation run as the phenotype of an individual worm. Thus, to reproduce a worm population, we performed 5000 simulation runs for each genetic background with different random seeds, each for 1000 maximally parallel steps.

Based on the current experimental knowledge (Shaye and Greenwald, 2002, 2005), we determine the fate adopted by each cell by measuring and correlating the concentration levels of MPK-1* and LIN-12*. Specifically, 1° fate is induced by a high level of MPK-1* and is refractory to LIN-12*. The 2° fate is induced by a low level of MPK-1* and a high level of LIN-12*. Low levels of both MPK-1* and LIN-12* lead to 3° fate. From the simulation, we calculate the LIN-12* and MPK-1* concentration levels as the average number of tokens over the final 50 steps to avoid unnecessary noise from the continual movement of tokens. Predicted cell fates are not influenced significantly by the length of the averaging because the protein concentration levels at the very end of the simulation are generally in a steady state. Assuming that a *high* concentration level corresponds to more

**Fig. 7.** Landscape produced by intersecting the three scoring functions. The quadrants are labelled with the corresponding cell fates.

than three tokens in a place and a *low* level corresponds three tokens or less, it is possible to determine the adopted fate. The corresponding piecewise function is formalized in the Supplementary Material.

To facilitate parameter adjustments during calibration, we also implemented three scoring functions (one for each cell fate) as sigmoids, in order to obtain a continuous curve instead of the discrete and discontinuous profile of a piecewise function. Such a continuous score was very useful during calibration to guide recovery of the expected behaviour by comparing slight changes in the scores produced by different adjustments.

Each scoring function rewards (i.e. score tends to 1) concentration levels that match the corresponding description. In the scoring functions, more than four tokens corresponds to *high* and less than two corresponds to *low*, while intermediate numbers of tokens (in between 4 and 2) produce the S-shaped gradient peculiar to sigmoids. In our experience, slight changes in the shape of the functions (e.g. steepness) do not significantly change the results. Consequently, each scoring function, using the simulated LIN-12* and MPK-1* concentration levels as variables, computes a score in the interval [0, 1] that measures how closely a simulated cell reproduces the fate description captured by the scoring function. For each cell, we calculate three scores (one for each function), and assign to the cell the fate corresponding to the function that returns the highest score. The analytic form of these functions can be found in the Supplementary Material.

The intersection of the three scoring functions generates a landscape (Fig. 7), which can be compared with the discrete representation of the piecewise function and resembles the fate plane proposed by Giurumescu *et al.* (2006), in which the quadrants identify cell fates.

3 RESULTS

3.1 Model validation

To determine the capability of our model to reproduce and predict the biological behaviour, we simulated 64 different experimental conditions. Twenty-two experiments (Table 2) previously selected in Fisher *et al.* (2005) were used for model calibration. Thirty perturbations were used for validation: 26 (see the Supplementary Material) from Fisher *et al.* (2005), three (Table 3) from Sternberg (2005), and one (Exp. 52, Table 4) from Yoo and Greenwald (2005). Particularly, experiment 51 (Table 3) was never simulated in any previous work that we are aware of. The remaining 12 simulations constitute new predictions. Of these, the most remarkable ones (Table 4) are discussed in Section 3.2. Statistical details for all simulations and a short animation displaying a typical single run are available in the Supplementary Material.

Our model reliably reproduces all the mutant combinations, except for the double mutant *lin-12(gf);lin-15(lf)* (Table 2, Exp. 21 and 45), even if in these cases, a fraction of the predictions matches the expected pattern. The noticeable differences of biological

Table 3. In vivo experiments not used for the model construction

Exp.	AC	Genotype		Fate pattern						Ref.
		let-60	lin-3	P3.p	P4.p	P5.p	P6.p	P7.p	P8.p	
49	+	lf		3	3	3	3	3	3	i
50	+		lf	3	3	3	3	3	3	j
51	+		gf	3	2	1	1	1	2	j

i: Beitel et al. (1990); j: Sternberg (2005).

Table 4. Selection of microRNA experiment outcomes predicted by our model. mir-61(ce) stands for constitutive expression of mir-61.

Exp.	AC	Genotype			Fate pattern						Ref.
		mir-61	Vul	lst	P3.p	P4.p	P5.p	P6.p	P7.p	P8.p	
52	+	ce			2	2	2	2	2	2	k
53	–	ce			2	2	2	2	2	2	
54	+	ce	lf		2	2	2	2	2	2	
55	+	ce		lf	2	2	1	1	1	2	
56	+	lf		lf	3	2	1	1	1	2	

k: Yoo and Greenwald (2005).

observations from different labs, and the few worms examined in vivo, do not help to establish a trustworthy expected outcome.

Of the 22 experiments in Table 2, particularly interesting are the experimental conditions that lead to unstable fate patterns. These results were already discussed in Fisher et al. (2007) and Sun and Hong (2007), but these discussions lacked statistical detail about the possible outcomes. In fact, Sun and Hong (2007) observed that the statecharts model of Fisher et al. (2005) often produces two adjacent 1° fate cells, which they claim is rarely observed in experiments, but they also do not provide supplementary statistical details.

In Table 5, we provide statistical details for Experiment 5 from Table 2. More than 93.4% of the predicted patterns match one of the expected biological 1½°-1½°-2°-1°-2°-1½° combinations. Of all matching patterns, only 4.5% contain three or more adjacent 2° fate cells, while just 2.7% have two or more adjacent 1° fate cells. These quantities correspond to the biological evidence that in these experiments three adjacent 2° fate, or two adjacent 1° fate cells are very unlikely. In the remaining 6.8% (not included in Table 5), one or more cells adopt 3° fate, and we interpret these outcomes as the ‘rare phenotypes’ in which uninterpretable lineages are observed (i.e. in between 2° and 3°), as noted for instance in Sternberg and Horvitz (1989).

In our approach, each maximally parallel step corresponds to a time step in the ontogeny of the biological system. Thus, our simulations can also be interpreted as time courses of gene regulation in vulval development. In Figure 8, the gene expression time series generated by our model are compared with the fluorescent photomicrographs published by Yoo et al. (2004). They show evidence of the graded expression of the egl-17p::cfp-lacZ reporter that responds to the Ras/MAPK pathway. Figure 8b depicts the time series generated by our model from the simulation results of a wild-type animal. Initially, MPK-1* (downstream product of the

Table 5. Detailed statistical results for the 5000 simulation of in vivo Experiment 5, Table 2 (lin-15(lf))

Exp.	Fate pattern						Occurrences	Percentage
	P3.p	P4.p	P5.p	P6.p	P7.p	P8.p		
5	Combinations matching the commonly observed pattern:						86.2	
	1	2	2	1	2	1	1348	27.0
	2	1	2	1	2	1	1180	23.6
	2	1	2	1	2	2	946	19.0
	1	2	2	1	2	2	830	16.6
	Three or more adjacent 2° fate cells:						4.5	
	2	2	2	1	2	1	132	2.6
	2	2	2	1	2	2	93	1.9
	Two adjacent 1° fate cells:						2.7	
	1	1	2	1	2	1	88	1.8
	1	1	2	1	2	2	46	0.9

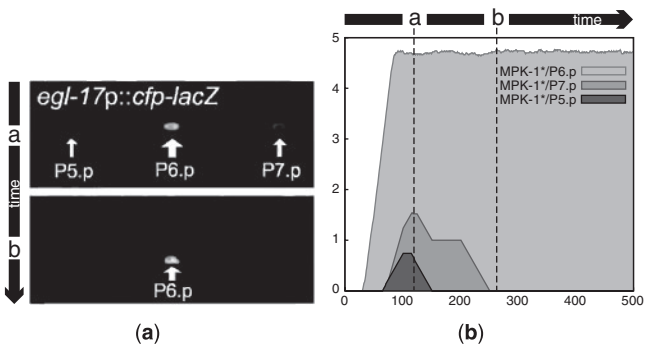


Fig. 8. Comparison between photomicrographs of gene activity by fluorescently labelled gene products, and simulation results. (a) Photomicrographs of the graded expression of the inductive signal adapted from Yoo et al. (2004), Science Magazine ©2004, AAAS. (b) Time series plot generated by our model, showing the graded expression of the inductive signal, initially faintly present in P5.p and P6.p. A running average over 50 steps is used for clarity of presentation. Concentration levels are on the vertical axis, while maximally parallel steps on the horizontal. One can correlate photomicrographs a and b with point a and b in the time series.

Ras/MAPK pathway as EGL-17) is faintly expressed in P5.p and P7.p. Subsequently, expression in P5.p and P7.p disappears, and MPK-1* remains at a high level only in P6.p, in accordance with the fluorescent photomicrographs of Figure 8a. We note that the concentration levels at the end of the simulation are approximately constant, indicating a steady state. In a related experiment, Yoo et al. (2004) divided lst genes into two groups: pattern A which contains dpy-23 and lst-3, and pattern B to which lst-1, lst-2 and lst-4 belong. Each group has its own characteristic temporal expression pattern that corresponds closely to the time series generated by our simulation (see the Supplementary Material).

3.2 mir-61: developmental switch and modulator

Our computational model, besides reproducing well-known biological experiments, encodes and unifies different published hypotheses and conjectures, shedding light on the vulval development process. The two hypotheses described next are

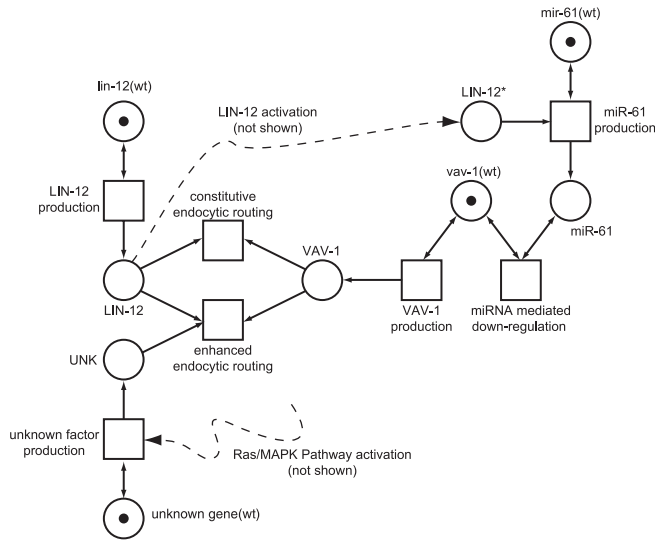


Fig. 9. Single model capturing different biological suggestions as explained in Section 3.2.

related to LIN-12 down-regulation, which is essential during vulval organogenesis (Shaye and Greenwald, 2002; Yoo *et al.*, 2004), and link the microRNA mir-61 to the vulva development process.

Shaye and Greenwald (2002) propose that, besides the degree of constitutive internalization displayed by LIN-12, Ras activation leads to transcription of an unknown factor that enhances the rate of internalization, promoting the endocytic routing of LIN-12. In Figure 9 one can see how we captured this hypothesis in our model. Activation of Ras enables the transcription of the unknown gene, which down-regulates LIN-12 post-translationally. Notably, changing the model of LIN-12 down-regulation from post- to pre-translation disrupts this behaviour and significantly alters our results.

Yoo and Greenwald (2005) identified mir-61 as direct transcriptional target of the LIN-12/Notch pathway. The gene mir-61 encodes a microRNA which blocks expression of the mRNA encoding VAV-1, a protein involved in LIN-12 down-regulation, possibly promoting LIN-12 endocytosis. They therefore proposed that activation of mir-61 by LIN-12 and the consequent down-regulation of VAV-1 constitute a positive-feedback loop that promotes LIN-12 activity in presumptive 2° fate VPCs. Although the unknown factor conjectured by Shaye and Greenwald does not seem to be required for the initial internalization of LIN-12, VAV-1 is necessary for the constitutive internalization of LIN-12. Notice that VAV-1 is involved in both constitutive and enhanced post-translation (endocytosis mediated) down-regulation of LIN-12.

Modelling these hypotheses (Fig 9) and capturing their behaviour has proven to be necessary to obtain the expected results during *in silico* experiments. Moreover, we simulated several perturbations of the mir-61 microRNA gene, obtaining the outcomes shown in Table 4. This appropriately confirms the role of the positive-feedback loop proposed by Yoo and Greenwald (2005). All experiments of Table 4, as far as we know, have not been tested *in vivo* (with the exception of experiment 52, which is described in Yoo and Greenwald, 2005).

Table 6. Detailed statistics for the simulation of Experiment 2 (*lst(lf)*), Table 2 and 56 (*mir-61(lf);lst(lf)*), Table 4

Exp.	Fate pattern						Occurrences	Percentage
	P3.p	P4.p	P5.p	P6.p	P7.p	P8.p		
2	3	3	1	1	1	3	4800	96.0
	3	3	1	1	1	2	199	4.0
56	3	2	1	1	1	2	1594	31.9
	3	3	2	1	1	2	1399	28.0
	3	3	2	1	2	3	1000	20.0
	3	2	1	1	2	3	998	20.0

Outcomes below 0.1% are omitted.

Experiments 52, 53, 54 and 55 confirm the specific role of mir-61 in influencing the cell fate decision, as determined by Yoo and Greenwald. Experiment 56 suggests a possible secondary role. This is a double mutant *mir-61(lf);lst(lf)* variation of the *lst(lf)* Experiment 2, Table 2. Although the single mutant *lst(lf)* expresses a stable VPC fate pattern, the loss-of-function of mir-61 in the double mutant disrupts the stability of the pattern, as can be seen in the statistical breakdown of Table 6. Based on this observation, we suggest that besides acting as developmental switch, mir-61 plays a ‘tuning’ role (Karp and Ambros, 2005) to ensure the stability of the cell fate pattern formation.

To the best of our knowledge, we are the first to model *in silico* microRNA interactions during *C.elegans* vulval induction, supporting the conjecture formulated in Yoo and Greenwald (2005), which lin-12, mir-61 and vav-1 form a feedback loop that helps to maximize lin-12 activity in the presumptive 2° VPCs.

4 DISCUSSION

Modelling and analysing developmental processes is a challenging task, as these biological processes often encompass several cells and evolve over the course of several hours. Moreover, the current lack of precise quantitative parameters at molecular level and the descriptive form of this biological knowledge welcome research on different modelling approaches able to reach the sweet spot in between abstraction and biological significance. In the work presented here, we abstracted the descriptive knowledge into a simple formal model that suitably mimics the underlying biological mechanisms and retains an adequate predictive power.

The Petri net used in our approach has a rather simple formalism, but the network designed by us is fairly large. Although several tools able to build extensive Petri nets with modular support exist (CPN Tools, 1999; Peccoud *et al.*, 2007), they are often quite complex in order to support much richer formalisms than the one we used, or they do not scale to the size of our Petri net model. Furthermore, the lack of a Petri net tool with a robust and efficient implementation of the maximal parallel execution semantics led us to build our own simulation tool (available on the web page of our project).

In conclusion, we applied our Petri net approach to *C.elegans* vulval development, reproducing several *in vivo* experiments. We generated insightful and testable predictions involving the microRNA mir-61. Our model is a suitable but partial representation of the whole intricate developmental process that leads to the

formation of the *C.elegans* vulva. New understanding of the process, supported by further experimental analysis, can be conveniently integrated in our model taking advantage of its modular fashion.

Funding: This work was supported in part by ENFIN, a Network of Excellence funded by the European Commission within its FP6 Program, under the thematic area “Life Sciences, genomics and biotechnology for health”, contract number LSHG-CT-2005-518254.

Conflict of Interest: none declared.

REFERENCES

- Beitel,G.J. et al. (1990) *Caenorhabditis elegans* ras gene let-60 acts as a switch in the pathway of vulval induction. *Nature*, **348**, 503–509.
- Berset,T. et al. (2001) Notch inhibition of RAS signaling through MAP kinase phosphatase LIP-1 during *C. elegans* vulval development. *Science*, **291**, 1055–1058.
- Berset,T.A. et al. (2005) The *C. elegans* homolog of the mammalian tumor suppressor Dep-1/Scc1 inhibits EGFR signaling to regulate binary cell fate decisions. *Genes Dev.*, **19**, 1328–1340.
- Burkhard,H.-D. (1980) On priorities of parallelism. In *Logics of Programs*, Vol. 148 of *LNCS*. Springer, London, UK, pp. 86–97.
- Chaouiya,C. (2007) Petri net modelling of biological networks. *Brief. Bioinform.*, **8**, 210–219.
- Chen,N. and Greenwald,I. (2004) The lateral signal for LIN-12/Notch in *C. elegans* vulval development comprises redundant secreted and transmembrane DSL proteins. *Dev. Cell*, **6**, 183–192.
- CPN Tools (1999) Available at <http://wiki.daimi.au.dk/cpntools/cpntools.wiki> (last accessed date October, 2008).
- Cui,M. et al. (1977) SynMuv genes redundantly inhibit lin-3/EGF expression to prevent inappropriate vulval induction in *C. elegans*. *Dev. Cell*, **10**, 667–672.
- Ferguson,E.L. and Horvitz,H.R. (1989) Post-embryonic cell lineages of the nematode, *Caenorhabditis elegans*. *Genetics*, **123**, 109.
- Fisher,J. et al. (2005) Computational insights into *Caenorhabditis elegans* vulval development. *Proc. Natl Acad. Sci. USA*, **102**, 1951–1956.
- Fisher,J. et al. (2007) Predictive modeling of signaling crosstalk during *C. elegans* vulval development. *PLoS Comput. Biol.*, **3**, e92.
- Fisher,J. et al. (2008) Bounded asynchrony: concurrency for modeling cell-cell interactions. In *Proceedings of the Formal Methods in Systems Biology 2008*, Vol. 5054 of *LNBI*, Springer, Cambridge, UK, pp. 17–32.
- Gilbert,D. et al. (2007) A unifying framework for modelling and analysing biochemical pathways using Petri nets. In *Proceedings of the Computational Methods in System Biology 2007*, Vol. 4695 of *LNCS*, Springer, Edinburgh, UK, pp. 200–216.
- Giurumescu,C.A. et al. (2006) Intercellular coupling amplifies fate segregation during *Caenorhabditis elegans* vulval development. *Proc. Natl Acad. Sci. USA*, **103**, 1331–1336.
- Goss,P. and Peccoud,J. (1998) Quantitative modeling of stochastic systems in molecular biology by using stochastic Petri nets. *Proc. Natl Acad. Sci. USA*, **95**, 6750–6755.
- Greenwald,I.S. et al. (1983) The lin-12 locus specifies cell fates in *Caenorhabditis elegans*. *Cell*, **32**, 435–444.
- Grunwald,S. et al. (2008) Petri net modelling of gene regulation of the Duchenne muscular dystrophy. *Biosystems*, **92**, 189–205.
- Gutenkunst,R.N. et al. (2007) Universally sloppy parameter sensitivities in systems biology models. *PLoS Comput. Biol.*, **3**, e189.
- Kam,N. et al. (2003) Formal modeling of *C.elegans* development: a scenario-based approach. In *Proceedings of the Computational Methods in System Biology 2003*, Vol. 2602 of *LNCS*. Springer, Trento, Italy.
- Karp,X. and Ambros,V. (2005) Encountering microRNAs in cell fate signaling. *Science*, **310**, 1288–1289.
- Kimble,J. (1981) Alterations in cell lineage following laser ablation of cells in the somatic gonad of *Caenorhabditis elegans*. *Dev. Biol.*, **87**, 286–300.
- Koch,I. and Heiner,M. (2008) Petri nets. ch. 7. In Junker,B.H. and Schreiber,F. (eds), *Analysis of Biological Networks*. Wiley-Interscience, New York, NY, pp. 139–180.
- Krepska,E. et al. (2008) Design issues for qualitative modelling of biological cells with Petri nets. In *Proceedings of the Formal Methods in Systems Biology 2008*, Vol. 5054 of *LNBI*, Springer, Cambridge, UK, pp. 48–62.
- Li,C. et al. (2009) Simulation-based model checking approach to cell fate specification during *Caenorhabditis elegans* vulval development by hybrid functional Petri net with extension. *BMC Syst. Biol.*, **3**, 42.
- Matsuno,H. et al. (2000) Hybrid Petri net representation of gene regulatory network. In *Proceedings of the Pacific Symposium on Biocomputing*, Vol. 5, World Scientific Publishing, Hawaii, USA, pp. 338–349.
- Matsuno,H. et al. (2006) Petri net based descriptions for systematic understanding of biological pathways. *IEICE Trans. Fundam. Electron. Commun. Comput. Sci.*, **E89-A**, 3166–3174.
- Peccoud,J. et al. (2007) Mobius: an integrated discrete-event modeling environment. *Bioinformatics*, **23**, 3412–3414.
- Peleg,M. et al. (2005) Using Petri net tools to study properties and dynamics of biological systems. *J. Am. Med. Inform. Assn.*, **12**, 181–199.
- Petri,C.A. (1962) *Kommunikation mit Automaten*. PhD Thesis, Technische Universität Darmstadt, Bonn, Germany.
- Reisig,W. and Rozenberg,G. (eds) (1998) *Lectures on Petri Nets I: Basic Models*, Vol. 1491 and 1492 of *LNCS*. Springer, Berlin, Germany.
- Sackmann,A. et al. (2006) Application of Petri net based analysis techniques to signal transduction pathways. *BMC Bioinformatics*, **7**, 482.
- Shaye,D.D. and Greenwald,I. (2002) Endocytosis-mediated downregulation of LIN-12/Notch upon Ras activation in *Caenorhabditis elegans*. *Nature*, **420**, 686–690.
- Shaye,D.D. and Greenwald,I. (2005) LIN-12/Notch trafficking and regulation of DSL ligand activity during vulval induction in *Caenorhabditis elegans*. *Development*, **132**, 5081–5092.
- Sternberg,P.W. and Horvitz,H.R. (1986) Pattern formation during vulval development in *C. elegans*. *Cell*, **44**, 761–772.
- Sternberg,P.W. and Horvitz,H.R. (1989) The combined action of two intercellular signaling pathways specifies three cell fates during vulval induction in *C. elegans*. *Cell*, **58**, 679–693.
- Sternberg,P.W. (2005) Vulval development. In *WormBook*. The *C. elegans* Research Community.
- Sulston,J.E. and Horvitz,H.R. (1977) Post-embryonic cell lineages of the nematode, *Caenorhabditis elegans*. *Dev. Biol.*, **56**, 110–156.
- Sundaram,M.V. (2004) Vulval development: the battle between Ras and Notch. *Curr. Biol.*, **14**, R311–R313.
- Sun,X. and Hong,P. (2007) Computational modeling of *Caenorhabditis elegans* vulval induction. *Bioinformatics*, **23**, i499–i507.
- Yoo,A.S. and Greenwald,I. (2005) LIN-12/Notch activation leads to microRNA-mediated down-regulation of Vav in *C. elegans*. *Science*, **310**, 1330–1333.
- Yoo,A.S. et al. (2004) Crosstalk between the EGFR and LIN-12/Notch pathways in *C.elegans* vulval development. *Science*, **303**, 663–666.