Bioinformatics, 2016, 1–5
doi: 10.1093/bioinformatics/btw299
Advance Access Publication Date: 10 May 2016
Original Paper

OXFORD

Systems biology

Computational discovery and *in vivo* validation of *hnf4* as a regulatory gene in planarian regeneration

Daniel Lobo[1], Junji Morokuma[2,3] and Michael Levin[2,3,*]

[1]Department of Biological Sciences, University of Maryland, Baltimore County, Baltimore, MD 21250, USA, [2]Center for Regenerative and Developmental Biology and [3]Department of Biology, Tufts University, Medford, MA 02155, USA

*To whom correspondence should be addressed.

Associate Editor: Jonathan Wren

Received on March 8, 2016; revised on April 25, 2016; accepted on May 4, 2016

Abstract

Motivation: Automated computational methods can infer dynamic regulatory network models directly from temporal and spatial experimental data, such as genetic perturbations and their resultant morphologies. Recently, a computational method was able to reverse-engineer the first mechanistic model of planarian regeneration that can recapitulate the main anterior–posterior patterning experiments published in the literature. Validating this comprehensive regulatory model via novel experiments that had not yet been performed would add in our understanding of the remarkable regeneration capacity of planarian worms and demonstrate the power of this automated methodology.

Results: Using the Michigan Molecular Interactions and STRING databases and the MoCha software tool, we characterized as *hnf4* an unknown regulatory gene predicted to exist by the reverse-engineered dynamic model of planarian regeneration. Then, we used the dynamic model to predict the morphological outcomes under different single and multiple knock-downs (RNA interference) of *hnf4* and its predicted gene pathway interactors *β-catenin* and *hh*. Interestingly, the model predicted that RNAi of *hnf4* would rescue the abnormal regenerated phenotype (tailless) of RNAi of *hh* in amputated trunk fragments. Finally, we validated these predictions *in vivo* by performing the same surgical and genetic experiments with planarian worms, obtaining the same phenotypic outcomes predicted by the reverse-engineered model.

Conclusion: These results suggest that *hnf4* is a regulatory gene in planarian regeneration, validate the computational predictions of the reverse-engineered dynamic model, and demonstrate the automated methodology for the discovery of novel genes, pathways and experimental phenotypes.

Contact: michael.levin@tufts.edu

1 Introduction

Planarian worms, despite their complex morphology, have the extraordinary ability to regenerate a complete body after almost any kind of amputation (Pearson and Sánchez Alvarado, 2008; Roberts-Galbraith and Newmark, 2015). With the goal to unravel the mechanisms responsible for this outstanding regenerative capacity, a large dataset of experiments, including surgical amputations, genetic perturbations and pharmacological interventions, and their resultant phenotypes have been reported in the literature and curated in the Planform centralized database (Lobo *et al.*, 2013a,b). However, due to the complex feed-back loops and non-linear interactions typical of biological regulatory networks, understanding how the worm

restore its head-trunk-tail pattern is a very difficult inverse problem (Lobo *et al.*, 2014). This has led to a lack of mechanistic, fully-specified models that can explain simultaneously more than one or two features of planarian regeneration (Lobo *et al.*, 2012; Umesono *et al.*, 2013).

In order to solve this inverse problem, heuristic computational methods have been proposed for the reverse-engineering of dynamic models directly from experimental data (Bonneau *et al.*, 2006; Schmidt and Lipson, 2009; Sirbu *et al.*, 2010; Yeung *et al.*, 2002), including biological development (Becker *et al.*, 2013; Crombach *et al.*, 2012; Jaeger *et al.*, 2004; Manu *et al.*, 2009; Perkins *et al.*, 2006; Reinitz *et al.*, 1995, 1998). These computational methods represent a promising approach to build systems-level models that can mechanistically explain the complex biological processes that are not only necessary but also sufficient for the observed exquisite regulation of growth and form.

Following this approach, the most comprehensive dynamic model of planarian regeneration was recently reverse-engineered with an automated computational method (Lobo and Levin, 2015). Directly from a dataset of outcomes of functional (surgical and genetic perturbation) experiments including the main head versus tail phenotypes, the method inferred a dynamic genetic regulatory network that could recapitulate all the experimental phenotypes when simulated *in silico*. Importantly, the model predicted the existence of unknown regulatory genes inferred as necessary for the correct regulation of planarian regeneration. Finding these unknown predicted products would aid in our understanding of planarian regeneration and serve as a validation of the reverse-engineered model in particular and the computational methodology approach in general.

We thus used the model to attempt to uncover new biological signaling and predict the outcomes of experiments that have never been performed. Here, we present the discovery of *hnf4* (hepatocyte nuclear factor 4) as a planarian regulatory gene and its predicted capacity by the reverse-engineered model to rescue an abnormal tailless phenotype. In addition, we tested these predictions *in vivo* by performing the same surgical and genetic experiments in the planarian worm, which validated all the predicted results by the reverse-engineered model.

2 Methods

2.1 Planarian culture

An asexual clonal line of *Schmidtea mediterranea* (CIW4) was maintained at 20°C in the dark, in 0.5 g/l Instant Ocean salts (Spectrum Brands, U.S.A.) added to ultrapure water. They were starved for at least a week before microinjection and double-stranded RNA interference (RNAi) experiments.

2.2 Double-stranded RNA interference (dsRNAi)

For double-stranded RNA interference experiments, dsRNA was generated and injected as described by Oviedo *et al.* (2008). dsRNA of each gene were microinjected into the worms of 7 mm in length, either as single gene dsRNAi, or mixed with another dsRNA for double-dsRNAi, while the amount of each dsRNA introduced was maintained constant. dsRNA injection was performed for 3 consecutive days, followed by amputation of the planaria into 3 regions (head, pharynx and tail) 5 h after the third and last injection. Amputations were performed under a dissecting microscope with a scalpel, on filter paper over a moist Kimwipe kept chilled on a cold plate. Pharynx fragments were produced by transversely cutting the worms immediately anterior and posterior of the pharynx. Fragments were kept in welled non-treated tissue-culture plates in the dark at 20 °C to regenerate, and moved into fresh media and new wells once a week, until scored.

2.3 Image collection, measurements and statistics

Images were collected using a Zeiss Stemi SV 6 dissecting microscope, with a Canon EOS Rebel T3i DSLR camera. All measurements were carried out using images of worms when fully extended. Adobe Photoshop was used on images of the regenerated pharyngeal fragment to categorize the pixels as belonging either to the head, trunk, or tail regions within the entire regenerating fragment and the pigment-less anterior and posterior blastemas. A custom script in Matlab was developed and used to automatically count the pixels in each region and calculate the ratios.

3 Results

3.1 Characterization of the predicted unknown regulatory gene as *hnf4*

The most comprehensive dynamic model of planarian regeneration, recently reverse-engineered with a computational method (Lobo and Levin, 2015), predicted an unknown regulatory gene necessary for the correct regeneration of the planarian worm (Fig. 1A, labeled with '?'). The model predicted that this unknown gene was regulated (directly or indirectly) by both *β-catenin* (MiMI gene id CTNNB1) and *hh* (hedgehog, MiMI gene id SHH). We then used this information to search for possible candidates in the Michigan Molecular Interactions database (MiMI) (Tarcea *et al.*, 2009), which resulted in the identification of *HNF4A* (MiMI gene id 3172) in *Homo sapiens* as the only known candidate to be regulated by both *β-catenin* (MiMI interaction id 48746) and *shh* (MiMI interaction id 228335) (Fig. 1B).

We further validated these interactions with the software tool MoCha (Lobo *et al.*, 2016). MoCha uses the STRING database v10 (Szklarczyk *et al.*, 2015) and facilitates the identification of pathways with unknown products. We queried MoCha for candidate products directly interacting with both *β-catenin* (ctnnb1) and *shh* (shh) in any organism and with no score limitations. The tool confirmed that *HNF4A* is a candidate for such regulations in *Homo sapiens*, *Mus musculus* and *Rattus norvegicus*. Table 1 shows the scores of the interactions found in the three species.

Interestingly, an homolog of *HNF4A* (*hnf4*) exists in the planarian *Schmidtea mediterranea*, and it has been found to be expressed in intestinal cells and in interspersed cells surrounding the intestine, including neoblasts (planarian stem cells) (Wagner *et al.*, 2011), to be involved in intestinal regeneration (Forsthoefel *et al.*, 2012), and to be expressed in the gamma neoblasts' subpopulation (van Wolfswinkel *et al.*, 2014). Furthermore, other homologs of *HNF4A* has been found to regulate blastema formation in amputated axolotl limbs (Jhamb *et al.*, 2011). Together, these data suggests that *hnf4* may be indeed a regulatory gene in planarian regeneration.

3.2 *In silico* morphological predictions of single and double knock-downs of *hnf4* and its interacting genes

We then used the reverse-engineered dynamic model of planarian regeneration to predict the morphological outcomes of amputated trunk pieces with single and double knock downs (RNAi) of *hnf4* and its predicted interacting genes *β-catenin* and *hh* (Fig. 1C). Consistent with the input dataset of experiments used to automatically infer the model, the control experiment resulted in the wild-type morphology, the knock down of *β-catenin* resulted in a double-head

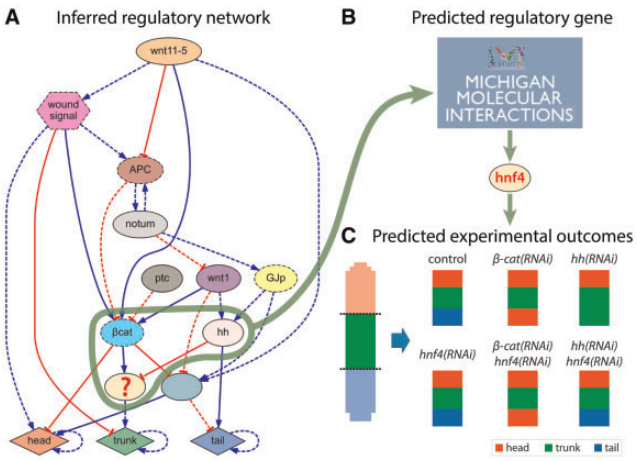

Fig. 1. Computational approach to discover a novel regulatory gene in planarian regeneration. (**A**) A reverse-engineered regulatory model from experimental data containing an unknown regulatory product (labeled '?'). (**B**) A search in the Michigan Molecular Interactions database for a product regulated by both *β*-catenin and hh in any organism returns hnf4 as the only candidate. (**C**) The reverse-engineered model predicts that knocking-down *hnf4* together with *hh* rescues the abnormal regenerated phenotype (tailless) resulting from knocking-down *hh* alone

Table 1. Characterization scores from the MoCha tool and the STRING database

Organism	MoCha score	Interactions	STRING score
Homo sapiens	138 346	CTNNB1→HNF4A	0.996
		SHH→HNF4A	0.474
Mus musculus	285 954	Ctnnb1→Hnf4a	0.930
		Shh→Hnf4a	0.247
Rattus norvegicus	285 954	Ctnnb1→Hnf4a	0.923
		Shh→Hnf4a	0.240

morphology (Gurley *et al.*, 2008; Iglesias *et al.*, 2008; Petersen and Reddien, 2008), and the knock down of *hh* resulted in a tailless morphology (Rink *et al.*, 2009).

Then, we used the model to formulate novel predictions regarding the predicted *hnf4* regulatory gene and its interactors. As shown in Figure 1C, the simulation of a single knock down of *hnf4* resulted in a wild-type morphology (similar to the control), a double knock down of *β-catenin* and *hnf4* resulted in a double-head morphology (similar to the single knock down of *β-catenin*), and, remarkably, a double knock down of *hh* and *hnf4* resulted in a wild-type morphology, instead of the tailless morphology of the single knock down of *hh*. In this way, the reverse-engineered model formulated the non-trivial prediction that RNAi of *hnf4* would rescue the tailless phenotype due to RNAi of *hh*.

3.3 Validation of the regulatory role of *hnf4* with single and double RNAi *in vivo* experiments

We then validated all the *in silico* predictions made by the reverse-engineered model with similar *in vivo* experiments in the planarian worm. Similarly to the simulations, trunk pieces were amputated under single and double RNAi knock downs (Fig. 2A). The *in vivo* results (Fig. 2B) showed that while RNAi of *hnf4* resulted in the regeneration of the wild-type morphology, RNAi of *β-catenin* resulted in a double-head morphology and RNAi of *hh* resulted in a tailless morphology, as it was predicted by the reverse-engineered model. Then, the *in vivo* experiments showed that a double RNAi of *β-catenin* and *hnf4* still resulted in the double-head morphology, but, remarkably, a double RNAi of *hh* and *hnf4* restored the wild-type

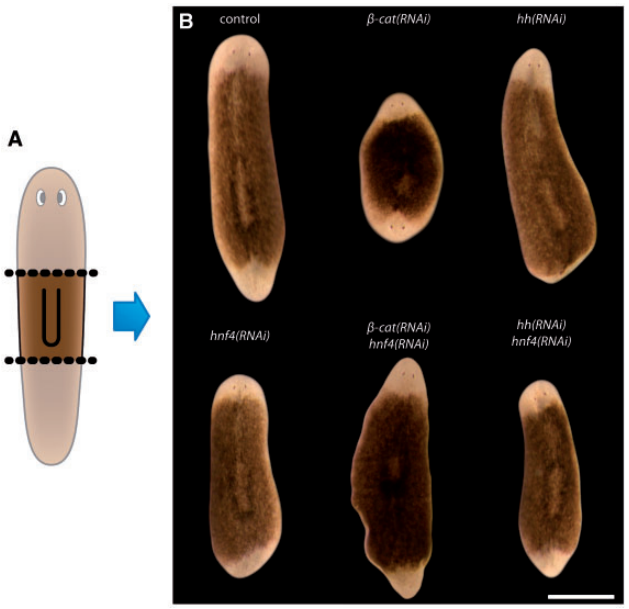

Fig. 2. Validation *in vivo* of the predicted phenotypes by the reverse-engineered model. (**A**) Diagram of the experimental surgical amputation of the head and tail regions in planarian worms. (**B**) Resulting phenotypes under different knock-down treatments, recapitulating all the predicted *in silico* morphologies. Original anterior facing up. Scale bar: 1 mm

morphology, instead of the tailless morphology obtained by the single RNAi of *hh*. These results validated the ability of RNAi of *hnf4* to rescue the tailless morphology, as it was predicted by the reverse-engineered model.

Additionally, we performed a statistical analysis of the area sizes of the regenerated tails under the different RNAi treatments (Fig. 3). The analysis confirmed that while there were no statistical difference in the tail area ratio between the control and the single RNAi of *hnf4*, the single RNAi of *hh* produced morphologies with statistically significant smaller tail area ratios. In the case of double RNAi of *hh* and *hnf4*, there were no statistical difference of the tail area ratios with respect the controls; however, the tail area ratios of the double RNAi morphologies were statistically larger than the single RNAi of *hh* morphologies. This statistical study validated the

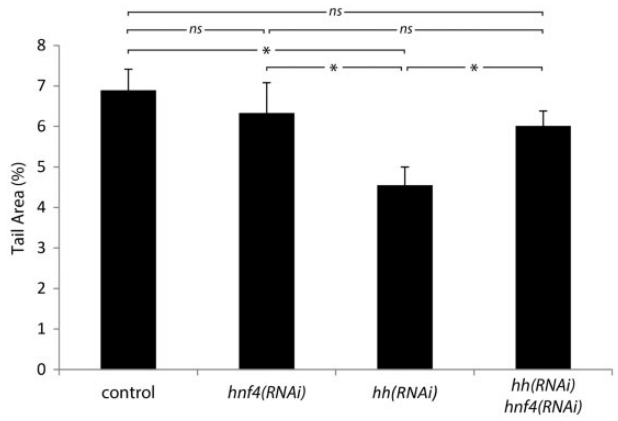

Fig. 3. Statistical analysis of the area size ratios of the regenerated phenotypes from a trunk piece under different gene knock-down treatments (RNAi). As predicted by the reverse-engineered model, the regenerated tail area ratio when knocking down *hh* is significantly smaller than the wild type but not significantly different than when knocking down both *hh* and *hnf4*. Each value is given as the mean ± SEM ($n = 24$–29). * $P < 0.05$, Student's *t*-test. *ns* not significant

predicted ability of RNAi of *hnf4* to rescue the tailless morphology due to RNAi of *hh*.

4 Discussion

Here, we presented the computational discovery and *in vivo* validation of *hnf4* as a regulatory gene in planarian regeneration. Starting from the prediction of an unknown necessary gene from a reverse-engineered dynamic model, we characterized it with the use of the MiMI and STRING databases and the MoCha tool. Then, we used the reverse-engineered dynamic model to produce exact predictions of the regenerated morphologies when perturbing *hnf4* and their regulatory interacting genes in the predicted pathway. This *in silico* study revealed the ability of RNAi of *hnf4* to rescue the abnormal regenerated phenotype (tailless) due to the RNAi of *hh*.

We then used an *in vivo* approach to validate at the bench all the predictions of the model. Using RNAi to knock-down single and double genes involved in the pathway, we performed similar surgical amputations and genetic knock downs (RNAi) in planarian worms. The results confirmed that RNAi of *hnf4* produced the wild-type morphology when applied alone, but when combined with RNAi of *hh*, it rescued the tailless phenotype caused by RNAi of *hh* alone. A statistical analysis further confirmed the significance of these results. In summary, these results show how the reverse-engineered model remarkably predicted a non-trivial regulatory role of *hnf4* in planarian regeneration.

The dynamic nature of the presented model explains in a mechanistic way the morphologies resulting from the different perturbations. When both *hh* and *hnf4* are knock-down, the level of tail signal is naturally higher than the trunk signal; that is, the default outcome is tail. When only *hnf4* is knock-down, the tail signal is even higher (since *hh* enhances tail) and the trunk signal stays the same than in the double knock-down, so the wild type morphology is still regenerated. However, when only *hh* is knock-down, *hnf4* is higher expressed (since *hh* represses *hnf4*), which results in a higher signal of trunk (since *hnf4* enhances trunk) and higher than tail (which is also lower than the wild type, since *hh* enhances tail), producing the no-tail morphology. Finally, with no knock-downs, the presence of *hh* produces higher signal of tail and reduces the signal of trunk (since *hh* enhances tail and represses *hnf4*, which enhances trunk), resulting in the wild type morphology.

5 Conclusion

Computational approaches are indispensable for the construction of predictive models of complex biological phenomena and the understanding of their regulatory networks. These methods can automatically formulate mathematical models that can recapitulate the observed phenotypes. Crucially, these reverse-engineered models can formulate testable predictions, including the prediction of novel regulatory elements and experimental phenotypes.

Here, we validated the novel predictions of a comprehensive dynamic model of planarian regeneration, including the discovery of a novel regulatory gene (*hnf4*) and its capacity to rescue an abnormal phenotype. The model accounts for the minimum set of interactors needed to explain the resultant head versus tail morphology from a comprehensive set of perturbation experiments. In a future work, we will validate the second unknown regulatory gene predicted by the model, which was preliminary characterized as the Frizzled family of receptors (Lobo and Levin, 2015). Further regulatory elements are indeed needed to control other aspects of tail regeneration, such as the specific round shape, determination of tissue types (epithelium, muscle, nerves, etc.) in the correct locations, and the dorsal-ventral axis regulation. Adding perturbation experiments involving these elements to the input dataset will allow for the discovery of more detailed models by the reverse-engineering system, extending the pathways presented here.

This work serves as a proof of principle for a pipeline for converting models discovered by machine learning into predictive hypotheses that can be tested at the bench to uncover novel biological interactions. Interaction databases such as MiMI or STRING enable the characterization of unknown elements predicted by these models. Importantly, the interaction confidence scores included in these databases summarize the quality of the available evidence and can be used to rank candidate interactions, being the ones with higher score the best candidate interactions to test at the bench. This approach demonstrates the strength of automated reverse-engineering methods for the discovery of predicted mechanistic models of developmental and regenerative biology.

Funding

This work has been supported by NSF grant EF-1124651. Computation used the Extreme Science and Engineering Discovery Environment (XSEDE), which is supported by NSF grant ACI-1053575, and a cluster computer awarded by Silicon Mechanics.

Conflict of Interest: none declared.

References

Becker,K. *et al.* (2013) Reverse-engineering post-transcriptional regulation of gap genes in *Drosophila melanogaster*. *PLoS Comput. Biol.*, **9**, e1003281.

Bonneau,R. *et al.* (2006) The Inferelator: an algorithm for learning parsimonious regulatory networks from systems-biology data sets de novo. *Genome Biol*, **7**, R36.

Crombach,A. *et al.* (2012) Efficient reverse-engineering of a developmental gene regulatory network. *PLoS Comput. Biol.*, **8**, e1002589.

Forsthoefel,D.J. *et al.* (2012) An RNAi screen reveals intestinal regulators of branching morphogenesis, differentiation, and stem cell proliferation in planarians. *Devel. Cell*, **23**, 691–704.

Gurley,K.A. *et al.* (2008) β-Catenin defines head versus tail identity during planarian regeneration and homeostasis. *Science*, **319**, 323–327.

Iglesias,M. *et al.* (2008) Silencing of Smed-βcatenin1 generates radial-like hypercephalized planarians. *Development*, **135**, 1215–1221.

Jaeger,J. *et al.* 2004) Dynamic control of positional information in the early Drosophila embryo. *Nature*, **430**, 371.

Jhamb,D. *et al.* (2011) Network based transcription factor analysis of regenerating axolotl limbs. *BMC Bioinformatics*, **12**, 80.

Lobo,D. and Levin,M. (2015) Inferring regulatory networks from experimental morphological phenotypes: a computational method reverse-engineers planarian regeneration. *PLoS Comput. Biol.*, **11**, e1004295.

Lobo,D. *et al.* (2012) Modeling planarian regeneration: a primer for reverse-engineering the worm. *PLoS Comput. Biol.*, **8**, e1002481.

Lobo,D. *et al.* (2013a) Planform: an application and database of graph-encoded planarian regenerative experiments. *Bioinformatics*, **29**, 1098–1100.

Lobo,D. *et al.* (2013b) Towards a bioinformatics of patterning: a computational approach to understanding regulative morphogenesis. *Biol. Open*, **2**, 156–169.

Lobo,D. *et al.* (2014) A linear-encoding model explains the variability of the target morphology in regeneration. *J. R. Soc. Interface*, **11**, 20130918.

Lobo,D. *et al.* (2016) MoCha: Molecular Characterization of Unknown Pathways. *J. Comput. Biol.*, **23**, 291–297.

Manu *et al.* (2009) Canalization of gene expression and domain shifts in the Drosophila blastoderm by dynamical attractors. *PLoS Comput. Biol.*, **5**, e1000303.

Oviedo,N. *et al.* (2008) Planarian PTEN homologs regulate stem cells and regeneration through TOR signaling. *Dis. Models Mechanisms*, **1**, 131–143.

Pearson,B.J. and Sánchez Alvarado,A. (2008) Regeneration, stem cells, and the evolution of tumor suppression. *Cold Spring Harbor Symposia Quant. Biol.*, **73**, 565–572.

Perkins,T.J. *et al.* (2006) Reverse engineering the gap gene network of *Drosophila melanogaster*. *PLoS Comput. Biol.*, **2**, 417–428.

Petersen,C.P. and Reddien,P.W. (2008) Smed-βcatenin-1 is required for anteroposterior blastema polarity in planarian regeneration. *Science*, **319**, 327–330.

Reinitz,J. *et al.* (1995) Model for cooperative control of positional information in Drosophila by bicoid and maternal hunchback. *J. Exp. Zool.*, **271**, 47–56.

Reinitz,J. *et al.* (1998) Stripe forming architecture of the gap gene system. *Dev. Genet.*, **23**, 11–27.

Rink,J.C. *et al.* (2009) Planarian Hh signaling regulates regeneration polarity and links Hh pathway evolution to cilia. *Science*, **326**, 1406–1410.

Roberts-Galbraith,R.H. and Newmark,P.A. (2015) On the organ trail: insights into organ regeneration in the planarian. *Curr. Opin. Genet. Devel.*, **32**, 37–46.

Schmidt,M. and Lipson,H. (2009) Distilling free-form natural laws from experimental data. *Science*, **324**, 81–85.

Sirbu,A. *et al.* (2010) Comparison of evolutionary algorithms in gene regulatory network model inference. *BMC Bioinformatics*, **11**, 59.

Szklarczyk,D. *et al.* (2015) STRING v10: protein–protein interaction networks, integrated over the tree of life. *Nucleic Acids Res.*, **43**, D447–D452.

Tarcea,V.G. *et al.* (2009) Michigan molecular interactions r2: from interacting proteins to pathways. *Nucleic Acids Res.*, **37**, D642–D646.

Umesono,Y. *et al.* (2013) The molecular logic for planarian regeneration along the anterior-posterior axis. *Nature*, **500**, 73–76.

van Wolfswinkel,J.C. *et al.* (2014) Single-cell analysis reveals functionally distinct classes within the planarian stem cell compartment. *Cell Stem Cell*, **15**, 326–339.

Wagner,D.E. *et al.* (2011) Clonogenic neoblasts are pluripotent adult stem cells that underlie planarian regeneration. *Science*, **332**, 811–816.

Yeung,M.K.S. *et al.* (2002) Reverse engineering gene networks using singular value decomposition and robust regression. *Proc. Natl. Acad. Sci. USA*, **99**, 6163–6168.