In Silico Identification of Endo16 Regulators in the Sea Urchin Endomesoderm Gene Regulatory Network

Huey-Eng Chua[§] Sourav S Bhowmick[§] Lisa Tucker-Kellogg[‡]
Qing Zhao[§] C F Dewey, Jr[†] Hanry Yu[¶]

[§]School of Computer Engineering, Nanyang Technological University, Singapore
[‡]Mechanobiology Institute, National University of Singapore, Singapore
[¶]Department of Physiology, National University of Singapore, Singapore
[†]Division of Biological Engineering, Massachusetts Institute of Technology, USA
chua0530|assourav|zhaoqing@ntu.edu.sg, LisaTK|nmiyuh@nus.edu.sg, cfdewey@mit.edu

ABSTRACT

Recent functional genomics research has yielded a large *in silico* gene regulatory network model (622 nodes) for endomesoderm development of sea urchin, a model organism for embryonic development. The size of this network makes it challenging to determine which genes are most responsible for a given biological effect. In this paper, we explore feasibility and accuracy of existing *in silico* techniques for identifying key genes that regulate `Endo16`, a widely-accepted gastrulation marker. We apply *target prioritization* tools (sensitivity analysis and PANI) to the endomesoderm network to identify key regulators of `Endo16` and validate the results by comparing against a set of benchmark `Endo16` regulators collated from literature survey. Our study reveals that global sensitivity analysis methods are prohibitively expensive and inappropriate for large networks. We show that PANI efficiently produces superior prioritization results compared to both random prioritization and local sensitivity analysis (LSA) techniques. Specifically, the area under the ROC curve was 0.625, ∼0.5, and 0.549 for PANI, random prioritization, and LSA, respectively. Our study reveals that certain unique characteristics of the endomesoderm network affect the performance of target prioritization techniques. In addition to identifying many known regulators of `Endo16`, PANI also discovered additional regulators (*e.g.*, `Snail`) that did not appear initially in the benchmark regulators set.

Categories and Subject Descriptors

J.3 [**Life and Medical Sciences**]: Biology and genetics.

General Terms

Algorithms, Experimentation, Performance, Verification

Keywords

PANI, sea urchin, endomesoderm, endo16, target prioritization

IHI'12, January 28–30, 2012, Miami, Florida, USA.

1. INTRODUCTION

Gastrulation is a process that happens early in embryogenesis when the blastula (unstructured assembly of cells) rearranges and forms the three germ layers (ectoderm, mesoderm, and endoderm) of the embryo [34]. These three germ layers subsequently differentiate and develop into different tissues and organs in the organogensis process. In the sea urchin, the gastrulation process consists of *primary* and *secondary invagination* [9]. In *primary invagination*, a portion of the epithelial wall of the blastula bend inwards creating the primitive gut known as archenteron. The *secondary invagination* starts when the archenteron has extended a distance of one-quarter to one-half across the blastocoel. Gastrulation defects can result in abnormal development of the body [14] and even death [29]. For instance, mutation of the `Shp2` phosphatase in zebrafish embryos result in convergence and extension cell movement defects. The phenotypes display craniofacial and cardiac defects similar to symptoms observed in human with Noonan and LEOPARD syndromes [14]. Although the gastrulation process varies across different organisms, there are certain characteristics which remain common. For instance, the gastrulation movements, such as invagination which is the inward bending of a sheet of cells, are preserved across species [36]. The use of animal models to study the gastrulation process enhances our understanding of the mechanisms underlying the developmental defects.

The use of monoclonal antibody and cloned gene probes enable the study of individual genes during gastrulation. `Endo16`, a cell surface glycoprotein, was first isolated from the purple sea urchin (*Strongylocentrotus purpuratus*) and characterized by [23]. The authors proposed that the `Endo16` protein may be involved in cell adhesion and gastrulation. Further studies in [30] identified `Endo16` as essential for gastrulation. Sea urchin *Lytechinus variegatus* embryos deficient in `Endo16` fail to undergo gastrulation and their blastocoele are filled with dissociated cells of unknown identity [30]. Understanding how regulators, such as `Otx`, affect `Endo16` protein expression brings us one step closer to discovering therapeutic targets for gastrulation defects.

1.1 Related Work and Motivation

Efforts in sea urchin developmental research have resulted in a vast accumulation of knowledge about different players in the gastrulation process [20, 23, 30]. In an attempt to in-

tegrate this knowledge, Davidson *et al.* [8] have constructed an ordinary differential equation (ODE) model describing the dynamic interactions between these different players based on experimental data from published literature. As the integrated network model grows large in size, it becomes increasingly difficult to study it manually. The endomesoderm gene regulatory network model described in [16] currently consists of 622 nodes (molecules) and 778 edges (interactions). In order to study the regulation of particular molecules (*e.g.*, `Endo16`), researchers have to sieve through the entire regulatory network to trace out relevant regulatory pathways. Hence, *in silico* techniques can play a key role in studying this problem by prioritizing the nodes that are likely to be relevant `Endo16` regulators. *However, to the best of our knowledge, no in silico study has been carried out to study the `Endo16` regulatory pathway in the sea urchin endomesoderm gene regulatory network.*

At first glance, it may seem that we can efficiently identify these target nodes by leveraging on the existing sensitivity analysis approaches [12, 27, 41]. Sensitivity analysis measures the effect of a parameter perturbation (*e.g.*, a kinetic rate constant change) on the node of interest, such as `Endo16`, and assigns sensitivity values to a node based on the extent of perturbation on `Endo16`. The parameter values of a real biological network vary due to differences in genetics, cellular environment and cell type. Hence, no single "true" nominal parameter value is deemed to exist. Thus, global sensitivity analysis (GSA) based methods, such as multi-parametric sensitivity analysis (MPSA) [41] and SOBOL [33], are deemed to be more appropriate for biological networks compared to local sensitivity analysis (LSA). GSA-based methods prioritize nodes using the sensitivity values when all parameters are varied simultaneously. These tools have been used widely to analyze several networks [41, 42]. However, our initial investigation revealed that these tools suffer from the following compelling limitations that prevent us from adopting them for investigating the endomesoderm gene regulatory network. First, they are computationally expensive as they require simulating the network behaviour for a combinatorial number of different parameter combinations. The use of GSA methods is limited to networks of smaller size. Particularly, both MPSA and SOBOL fail to perform the study of `Endo16` regulators in the large endomesoderm network on a modern server machine due to memory issues[1]. Second, prioritization based only on the sensitivity values means that "insensitive" nodes that may be important regulators may be missed. Lastly, as we shall see in Section 2.1, the sea urchin endomesoderm network is *partially correct* or *partially complete*. Unfortunately, sensitivity analysis based approaches are not robust enough to generate robust results from such networks. In summary, the aforementioned limitations have been the key obstacles for the research community to undertake systematic *in silico* strategy to study the `Endo16` regulatory pathway in the endomesoderm gene regulatory network.

1.2 Overview

This paper takes a first step to investigate the use of *in silico* target prioritization tools to identify regulatory nodes of `Endo16` in the endomesoderm gene regulatory network. *Target prioritization* is the problem of choosing a set of regulatory molecules specific to a particular node of interest (output node) that is related to the biological problem under investigation [7]. In this work, we chose `Endo16` as the output node for the endomesoderm network due to its critical role in gastrulation.

Recently, we proposed a generic algorithm for target prioritization called PANI (**P**utative T**A**rget **N**odes Pr**I**oritization), which uses network information and simple empirical scores to prioritize and rank biologically relevant target molecules in signaling networks [7]. PANI takes a two-phase approach to identify and rank target molecules. First, it prunes the nodes based on a reachability rule to eliminate nodes that are likely to be non-regulators. Then, it calculates the *putative target score* of each resulting node, which is a weighted rank aggregation of a dynamic property (*profile shape similarity distance* (PSSD)) and two structural properties (*target downstream effect* (TDE) and *bridging centrality* (BC) [13]) of the node. In [7], we demonstrated that PANI can prioritize a majority of drug targets that regulate `Erk` in the `MAPK-PI3K` network (containing only 36 nodes). Furthermore, the quality of results generated by this approach is superior to the GSA-based techniques. Hence, in this paper we investigate whether PANI can also be exploited to prioritize targets specific to `Endo16` regulation in the large endomesoderm network containing more than 600 nodes.

Our study reveals several interesting findings. PANI is successful in producing superior quality results by prioritizing many known `Endo16` key regulators in around 250 seconds on a modern desktop machine. We also observe that the endomesoderm network has certain unique structural and dynamic characteristics. Specifically, it contains a very large *strongly connected component* (SCC) and many nodes have constant concentrations. Consequently, the structural properties (*e.g.*, BC) in PANI play a more critical role compared to the dynamic property (PSSD) in producing superior quality results compared to random prioritization and LSA, which are oblivious to these characteristics. Note that PANI provides us the flexibility to tune the relative *weights* of structural and dynamic properties according to the characteristics of the underlying network. Lastly, PANI identified several target molecules (*e.g.*, `Snail`) that were not initially part of the set of benchmark regulators which we harnessed during literature survey. Further investigation revealed that these molecules indeed play a role in regulating `Endo16`. Hence, in addition to identifying many known regulators of `Endo16`, PANI's prioritization results give us a clue to additional targets that may also be regulators.

The rest of the paper is organized as follows. In Section 2, we describe the sea urchin endomesoderm gene regulatory network model used for analysis. In Section 3, we describe the use of PANI to prioritize the `Endo16` regulators and the steps to validate the results. PANI's prioritization results are then presented and discussed in Section 4. We discuss how PANI's parameters affect the result quality in Section 5.

2. ENDOMESODERM NETWORK

In this section, we summarize the general characteristics of the endomesoderm gene regulatory network model and briefly describe the biological process (endomesoderm specification) described by this network. The `Endo16` regulatory pathway (Figure 1a) which we use to validate our results in Section 4 forms a portion of this network. We create the

[1] SBML-SAT is used to perform MPSA and SOBOL analysis and is obtained from http://sysbio.molgen.mpg.de/SBML-SAT/. The default number of simulations is set to 2000 and 10000 for MPSA and SOBOL, respectively.

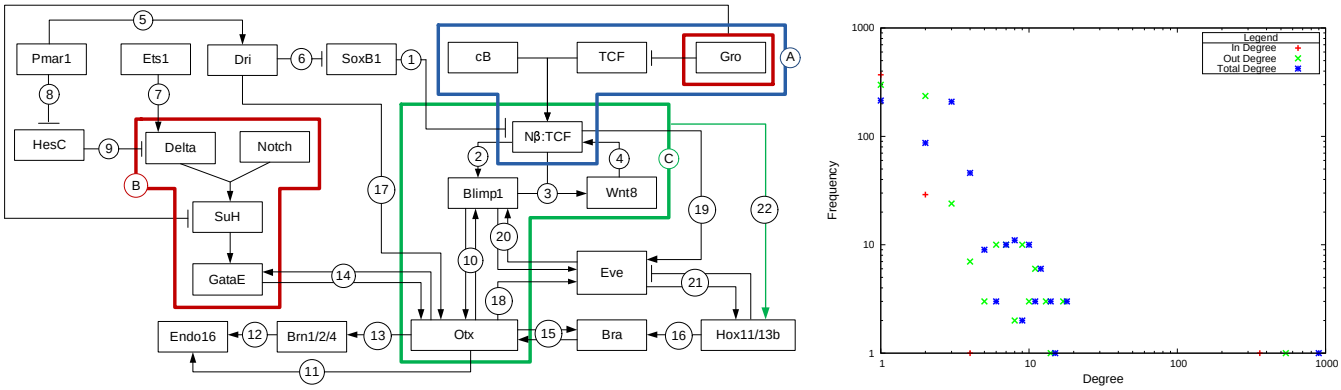

Figure 1: (a) The sea urchin `Endo16` regulatory pathway. Edges and modules (blue, red and green boxes) are labelled and elaborated in Section 2.2 and (b) Degree distribution of the endomesoderm network.

`Endo16` regulatory pathway based on literature survey and the scope of the survey is described in Section 3.2.

2.1 Network Characteristics

We obtain the ODE model of sea urchin endomesoderm gene regulatory network (BIOMD0000000235) from the *Biomodels.net* database [18]. This model is constructed from numerous perturbation experiments and contains 622 nodes and 778 edges. The nodes consist of 217 root nodes (with no incoming edges), 4 singletons (nodes with no incoming or outgoing edges) and 401 intermediate nodes (with incoming and outgoing edges). Amongst the intermediate nodes, 25 are not in any SCCs[2]. Of the remaining intermediate nodes, there are 8 SCCs containing two nodes and one huge SCC containing 360 nodes. The high percentage of nodes ($\sim 60\%$) involved in SCCs implies that many of the molecules are involved in autoregulation (a molecule regulating its own activity), a characteristic common in gene regulatory networks [19]. Figure 1b shows that the degree distribution of the endomesoderm network follows the power-law. Another characteristic of *in silico* models is their incompleteness, which may be due to missing genes or interactions [16], or to the approach used for model construction. In the case of the endomesoderm network, the authors use a heuristic-based approach to construct the network kinetics as it is impractical to perform parameter estimation for the entire network due to its large size. Validation against a similar subnetwork constructed using parameter estimation shows that there is 74% agreement of the simulation results. When compared to experimental data, the level of agreement falls to 48%. *Hence, the endomesoderm network is partially correct.* Note that such partial correctness is a real-world feature of many biological networks. Hence, any *in silico* approach for prioritizing biologically relevant targets must be robust enough to handle such networks.

2.2 Endomesoderm Specification

The network model used in [16] is an extension of that proposed in [8]. Although the model is partially correct (Section 2.1), it is still able to describe the key steps in endomesoderm development, namely, the initiation of the endomesoderm specification signal, the maintenance of the specification signal, the activation of the `Delta/Notch` signaling pathway, and the specification of $\texttt{veg}_1$ endoderm. Hence, it is still useful for our *in silico* study of the regulation of `Endo16`, whose expression is one of the crucial end points of the endomesoderm specification. In subsequent description, annotations of edges and modules (blue, red and green boxes) refer to that in Figure 1a.

Initiation of the endomesoderm specification signal. The single-cell zygote undergoes cleavage to form a multicell embryo. By the 6^{th} cleavage, the initial specification of the $\texttt{veg}_2$ domain occurs. This step requires two inputs: an intracellular signal from the micromeres and the nuclearization of β`-catenin` (`cB`), a cofactor required by the `TCF` transcription regulator for gene activation [8]. The nuclearization of `cB` relieves the repression of `TCF` by `Groucho` (`Gro`) as `TCF` binds with the nuclearized `cB` (`n`β) to form a complex (`n`β`:TCF`) (blue box A) [8]. `n`β`:TCF` activates several genes, including `Blimp1` [4].

Maintenance of the specification signal. The activity of `n`β`:TCF` is regulated positively by a feedback loop involving `Blimp1` and `Wnt8`; and negatively by its repressor, `SoxB1` [2] (edge 1). In the feedback loop, `n`β`:TCF` activates `Blimp1` (edge 2) and together with `Blimp1` results in the activation of `Wnt8` (edge 3) [8]. This in turn initiates the amplification of the endomesoderm specification activation signals (edge 4) [8]. `Dri` which is positively regulated by `Pmar1` [24] (edge 5), affects the late vegetal clearance of `SoxB1` in the $\texttt{veg}_2$ domain [1] (edge 6). The `n`β`:TCF` signal is required for expression of many $\texttt{veg}_2$ endomesodermal regulatory genes in the early to mid blastula stage, such as `Gcm` [8].

Activation of the `Delta/Notch` signaling pathway. At around the 8^{th} to 9^{th} cleavage, the micromeres express `Delta`, a ligand which activates the `Notch` receptor in the $\texttt{veg}_2$ cells, thus initiating the specification of these cells as mesodermal precursors [24]. Genes under the control of the `Notch` pathway, such as `GataE`, are expressed [11]. The `Delta/Notch` signaling is effected by the `Suppressor of Hairless` (`SuH`) transcription factor which is initially inhibited by `Groucho` (`Gro`) (red box B) [22]. The activity of `Delta` is in turn modulated by several molecules, namely, `Ets1`, `HesC` and `Pmar1`. `Ets1` has been implicated in downregulation of `Delta` expression at the late blastula stage when `Erk` is inhibited [31]. Inhibition of `Erk` prevents the phosphorylation of `Ets1` on `Thr107`, thus inhibiting `Ets1` (edge 7) [31]. `HesC-Pmar1` provides a double-negative control of `Delta` activity, whereby

[2] In a given SCC containing nodes u and v, there exists a path from u to v and vice versa.

Pmar1 inhibits HesC activity (edge 8) which in turn, inhibits Delta activity (edge 9) [6]. Hence, Pmar1 and Ets1 activate Delta while HesC inhibits Delta. The Neutralized-like-1 (Nrl) homolog in *Drosophila*, Neuralized (Neur), acts as a ubiquitin ligase which promotes the internalization and degradation of Delta [17], suggesting that Nrl may interact with Delta in the sea urchin in the same way. However, no supporting evidence has yet been found in the sea urchin. Hence, we did not consider Nrl as part of the Endo16 regulatory pathway in Figure 1a.

Specification of $\mathtt{veg}_1$ endoderm. At the late blastula stage, specification of the $\mathtt{veg}_1$ endoderm takes place. In this step, endodermal markers such as Endo16 are expressed [8]. Initially, Endo16 is expressed in the vegetal plate of the blastula [30]. The expression of Endo16 is regulated differently depending on the cell type and the embryogenesis phase. For instance, in primary mesenchymal cells (PMC), expression of Endo16 is downregulated [30]; Endo16 expression is maintained throughout the invaginating archenteron during gastrulation but downregulated in the anterior one-third of the archenteron at the end of gastrulation [30]. Specifically, the expression of Endo16 is regulated by Blimp1, Otx and Brain-1, -2, and -4 (Brn1/2/4). The initial activation of Endo16 in the endomesoderm is a result of Blimp1 activation of Otx (edges 10 and 11) [20, 38], while the late phase expression of Endo16 is regulated by Brn1/2/4 [40]. In [40], morpholino-substituted antisense oligonucleotide (MASO) treatment depresses the expression of Endo16 Module B significantly (edge 12). Quantitative PCR (QPCR) perturbation data at the later endoderm stage suggests that Otx drives the expression of Brn1/2/4 (edge 13) [40].

The activity of Otx is in turn regulated by several molecules, namely, Blimp1, GataE, Bra, Hox11/13b and Dri. There are three positive feedback loops that maintain Otx activity. The first two loops involve Blimp1 (edge 10) and GataE (edge 14) which interact with the β1/2 transcription unit of Otx [39]. The third loop involves Bra and Hox11/13b. Bra, a target gene of Otx, activates Otx (edge 15). The Bra-induced amplification of Otx is further amplified by Hox11/13b activation of Bra (edge 16) [28]. Dri is found to positively regulate the activity of Otxβ1/2 from QPCR perturbation data [1] (edge 17).

Another player in the Endo16 regulatory pathway is Even-skipped (Eve). Experimental data in [32] shows that Eve is regulated by four other molecules, namely, Otx, Blimp1, Hox11/13b and nβ:TCF. Both Otx (edge 18) and nβ:TCF (edge 19) activate Eve. The remaining nodes, Blimp1 and Hox11/13b, form a separate autoregulatory loop with Eve. In the Blimp1/Eve loop, both Blimp1 and Eve are positively activated (edge 20); in the Hox11/13b/Eve loop, Eve is repressed while Hox11/13b is activated (edge 21). Observation of the spatial expression of Hox11/13b in the vegetal plate in [3] suggests that Hox11/13b is downstream of the Wnt8/Blimp1/Otx (green box C) positive autoregulatory loop (edge 22).

Interested readers may refer to [8] and [24] for a detailed description of the model.

3. IN SILICO PRIORITIZATION

In this section, we describe our approach to identify and prioritize Endo16 regulators in the endomesoderm network. Our approach consists of two key steps. First, target prioritization was performed by exploiting the algorithm PANI [7]. Second, the results generated by the previous step were validated. Target prioritization and all subsequent experiments were carried out on an Intel 1.86GHz dual core processor machine with 2GB RAM, running Microsoft Windows XP.

3.1 Step 1: In Silico Prioritization

PANI [7] is a generic target prioritization algorithm that suggests target proteins for drug development, by predicting the most influential nodes in a disease-related signaling network. We have chosen to apply PANI (a two-phase algorithm described in [7]) to the problem of identifying key regulators of gastrulation. Briefly, the first (pruning) phase of PANI tests for the existence of a path between each node in the endomesoderm network and the node of interest, Endo16. Nodes having such paths are retained for further analysis in the next phase. Specifically, at the end of the first phase, 606 nodes are selected for subsequent processing. In the second phase (prioritization phase), a *putative target score* is calculated for each node and used for prioritization. The *putative target score* is a weighted rank aggregation of the *profile shape similarity distance* (PSSD), the *target downstream effect* (TDE) and the *bridging centrality* (BC) [13] of the nodes, which we elaborate in turn.

The first property, PSSD, identifies the most relevant upstream regulators of Endo16 by assessing the similarity between the concentration-time series profile (plot of a node's concentration against time) of each node with that of Endo16. Specifically, the PSSD between Endo16 and node v is calculated as the minimum dynamic time warping (DTW) distance [15] between two pairs of concentration-time profiles, namely $\{\zeta_{\mathtt{Endo16}},\zeta_v\}$ and $\{\zeta_{\mathtt{Endo16}},\zeta'_v\}$ where ζ'_v is the inverted profile of node v. In this paper, the concentration-time profiles are obtained from *in silico* simulations of the endomesoderm network model [16] using *Copasi* with parameters: {*duration*=70 hours, *intervals*=0.1 hours}[3]. That is, the length of the concentration time series ($|\zeta|$) is set to 700. The second property, TDE, measures the potential impact on the network when a node is perturbed. It is calculated as the sum of the effect of each of its downstream node w, which is the product of w's degree and the probability of perturbing w. The probability of perturbing w depends on the likelihood of the existence of a path leading to w. In the case of the endomesoderm network, we set this probability as 1 since the network is constructed based on extensive literature survey [16]. The last property, BC, identifies nodes that are located at a connecting bridge between modular subregions in a network [13]. It is calculated as the product of two ranks, namely, the inverses of *betweenness centrality* [5] and *bridging coefficient* [13].

The choice of the relative weights for the aforementioned properties in order to compute putative target score is influenced by the topological and dynamic characteristics of the network. For instance, PANI's computation of the PSSD ranks depend on similarity of changes in the concentration-time series profiles [7]. Consequently, the presence of many nodes having constant profiles in the network affects the PSSD rank and hence the prioritization results. Interestingly, in the endomesoderm network, 49.2% of the 197 nodes related to the benchmark regulators have constant profiles. Additionally, the presence of a large SCC in the network also

[3]The simulation time is unrelated to the duration parameter which intuitively, corresponds to the range of ζ and is related to $|\zeta|$ ($\frac{duration}{interval} = |\zeta|$).

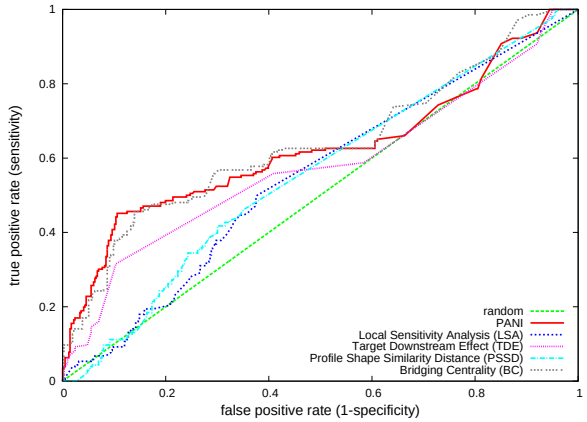

Figure 2: ROC curve.

have an effect on TDE and BC rankings. Hence, we allocate relatively lesser weights to PSSD and TDE compared to BC. Specifically, we set ω_{PSSD}=0.1, ω_{TDE}=0.2 and $\omega_{\frac{1}{\text{BC}}}$=0.7. Note that this is in contrast to the weights of these properties for the MAPK-PI3K network where PSSD is given higher weightage compared to TDE and BC [7]. In Section 5, we shall investigate the effect of different values of these weights on the target prioritization for the endomesoderm network.

3.2 Step 2: Validation of Results

A key issue of the previous step is the validation of the quality of the prioritization results. The purpose of the target prioritization is to identify the key regulators of Endo16. Hence, we will evaluate the quality of the results in terms of the biological relevance of the prioritized targets as Endo16 regulators in the sea urchin endomesoderm network.

We collate a list of known sea urchin Endo16 regulators based on extensive literature survey and use it as benchmark for validating the prioritized targets. We note that the use of literature survey for validation of biological relevance has limitations. For instance, the result of the validation is affected by the literature survey process, such as the keywords and selection criteria used for gathering and selecting the relevant literature. In order to keep our survey process as relevant to the problem as possible, we have looked for literature pertaining specifically to the sea urchin endomesoderm specification. We used the keywords "sea urchin endomesoderm" to search the *PubMed* repository and 73 publications were returned as of July 1, 2011. The literature survey was done based on these publications.

The specific steps for results validation were as follows.

- First, we constructed the sea urchin Endo16 regulatory pathway by mapping out the interactions between different molecules from the publications. We restricted our regulatory pathway (Figure 1a) to reflect nodes in the network model [16] that were relevant to Endo16 regulation to facilitate our validation later, as nodes in this regulatory pathway would be used as the benchmark set of Endo16 regulators. This benchmark set of regulators were made up of 20 different molecules. These molecules were represented as multiple nodes in the network (Tables 1 and 2), each of which is a different form of the molecule (*e.g.*, protein, gene, mRNA) in different embryonic territories (*e.g.*, endoderm, mesoderm and primary mesenchyme cells (PMC)). For instance, Protein P Otx is the Otx protein in the PMC.
- Next, we evaluated the quality of the results by assessing how well the top ranking (top 10%) nodes correspond with the benchmark set of Endo16 regulators. We also evaluated the sensitivity and specificity of our prioritization technique to identify the set of Endo16 regulators using Receiver Operating Characteristic (ROC) analysis (Section 4.1).
- Finally, we compared the performance of PANI with random prioritization and local sensitivity analysis (LSA) in the context of the endomesoderm network (Section 4.2).

4. VALIDATION OF IDENTIFIED TARGETS

In this section, we validate the quality of our prioritized results. In order to assess how well the prioritization results can identify the set of benchmark regulators, we examine the correlation between the list of PANI's top ranking nodes and the benchmark regulators, and perform a ROC analysis. For a more complete analysis, we also compare PANI to two baseline approaches, namely random prioritization and the LSA-based approach.

4.1 Top-Ranked Nodes and ROC Analysis

From the prioritized list of targets generated by PANI, we take the top 10% of nodes to test for enrichment of known Endo16 regulators. Tables 1 and 2 report the ranks of all the nodes based on their putative target scores. Recall from Section 3.1 that the size of the pruned set of candidate nodes is 606. Hence, there are 61 nodes in the top 10%. Observe that the top 61 nodes in Tables 1 and 2 consist of 25 different molecules V = {Wnt8‡, Bra‡, Hox‡, cB‡, Delta‡, GataE‡, Notch‡, Otx‡, Pmar1‡, SoxB1‡, Ets1‡, HesC‡, Dri‡, Erg, Hex, Hnf6, Snail, Tgif, VEGFR, SoxC, Tel, VEGFSignal, Sm30, Gcm, Gcad} as some of these molecules are represented multiple times. Molecules marked with ‡ are implicated in the endomesoderm specification process that controls Endo16 activity (Section 2.2) and represent a significant percentage in the top 61 nodes. For instance, cB is represented as Protein M cB, Protein E cB and Protein P cB, referring to β-catenin protein in the endoderm, in the mesoderm and in primary mesenchyme cells (PMC), respectively. *In total, 45 (74%) of the 61 putative target nodes are implicated in the regulation of Endo16, implying that the top 10% nodes are enriched with known Endo16 regulators.*

We note that 12 molecules in V not marked with ‡ do not correspond with the benchmark regulators in Figure 1a. We extended our literature search beyond the 73 publications to look for evidence implicating these molecules {Gcm, Hnf6, Tgif, Erg, Hex, Snail, Gcad, VEGFR, SoxC, Tel, VEGFSignal, Sm30} in the Endo16 regulatory pathway. We found that GataC is activated by Gcm [11] and Hnf6 [26] and indirectly inhibited by Alx1. Knockdown of GataC correlates strongly with down-regulation of FoxA [16] which inhibits Bra‡ [10]; Alx1 is activated by Tgif which is also involved in positive double feedback loops containing Erg and Hex [25]; Snail represses Gcad activity [37] which plays an inhibitory role on the nuclearization of cB‡ [21]. Hence, many of these nodes regulates the benchmark Endo16 regulators either directly or indirectly. Only VEGFR, SoxC, Tel, VEGFSignal, and Sm30 were not found linked with the benchmark regulators. *PANI's prioritization results identify both benchmark Endo16 regulators and additional nodes that are likely to play a regulatory role.*

ID	Node Name	Ψ_P	Ψ_L	ID	Node Name	Ψ_P	Ψ_L	ID	Node Name	Ψ_P	Ψ_L	ID	Node Name	Ψ_P	Ψ_L	ID	Node Name	Ψ_P	Ψ_L
1	GENE E Alx1	306‡	126	90	*GENE M Otx	310‡	126	179	mRNA E FvMo	162‡	68	268	mRNA M Tel	187‡	126	357	PRE M UMANrl	316‡	126
2	GENE E Apobec	308‡	126	91	GENE M Pks	309‡	126	180	mRNA E GataC	214	68‡	269	mRNA M Tgif	258‡	126	358	PRE M UMR	316‡	126
3	*GENE E Blimp1	313‡	126	92	*GENE M Pmar1	309‡	126	181	*mRNA E GataE	114‡	84	270	*mRNA M UbiqSoxB1	154	5‡	359	*PRE P cB	316‡	126
4	*GENE E Bra	310‡	126	93	GENE M Sm27	313‡	126	182	mRNA E Gcad	25	5‡	271	*mRNA M UMADelta	281	102‡	360	*PRE P Ets1	315‡	126
5	*GENE E Brn	305‡	126	94	GENE M Sm30	307‡	126	183	mRNA E Gcm	61‡	51	272	mRNA M UMANrl	280	103‡	361	PRE P Gcad	315‡	126
6	GENE E CAPK	318	126	95	GENE M Sm50	314‡	126	184	mRNA E Gelsolin	107‡	67	273	mRNA M UMR	281	102‡	362	PRE P L1	316‡	126
7	GENE E CyP	309‡	126	96	GENE M Snail	307‡	126	185	*mRNA E HesC	167‡	126	274	mRNA M VEGFR	247‡	126	363	*PRE P Otx	315‡	126
8	*GENE E Delta	310‡	126	97	*GENE M SoxB1	308‡	126	186	mRNA E Hex	250‡	109	275	*mRNA M Wnt8	42‡	61	364	PRE P UbiqAlx1	316‡	126
9	GENE E Dpt	307‡	126	98	GENE M SoxC	309‡	126	187	mRNA E Hnf6	193‡	126	276	mRNA M z13	203‡	126	365	PRE P UbiqES	316‡	126
10	*GENE E Dri	308‡	126	99	GENE M SuTx	308‡	126	188	*mRNA E Hox	73‡	70	277	mRNA P Alx1	146	37‡	366	*PRE P UbiqEts1	316‡	126
11	GENE E Endo16	308‡	126	100	GENE M TBR	309‡	126	189	mRNA E Kakapo	107‡	67	278	mRNA P Apobec	226	62‡	367	*PRE P UbiqHesC	316‡	126
12	GENE E Erg	309‡	126	101	GENE M Tel	309‡	126	190	mRNA E Lim	178‡	77	279	*mRNA P Blimp1	205	78‡	368	PRE P UbiqHnf6	316‡	126
13	GENE E ES	309‡	126	102	GENE M Tgif	310‡	126	191	mRNA E Msp130	262‡	126	280	*mRNA P Bra	115‡	77	369	PRE P UbiqSoxC	316‡	126
14	*GENE E Ets1	308‡	126	103	GENE M VEGFR	310‡	126	192	mRNA E MspL	227‡	126	281	*mRNA P Brn	273	68‡	370	PRE P UbiqTel	316‡	126
15	*GENE E Eve	310‡	126	104	*GENE M Wnt8	309‡	126	193	mRNA E Not	249	68‡	282	mRNA P CAPK	307‡	126	371	PROTEIN E Alx1	237‡	116
16	GENE E Ficolin	310‡	126	105	GENE M z13	309‡	126	194	*mRNA E Notch	281	101‡	283	*mRNA P cB	152	5‡	372	PROTEIN E Apobec	286	35‡
17	GENE E FoxA	313‡	126	106	GENE P Alx1	311‡	126	195	mRNA E Nrl	195‡	126	284	mRNA P CyP	117‡	77	373	*PROTEIN E Blimp1	127‡	113
18	GENE E FoxB	311‡	126	107	GENE P Apobec	308‡	126	196	mRNA E OrCt	224	62‡	285	*mRNA P Delta	58‡	77	374	*PROTEIN E Bra	4‡	126
19	GENE E FoxN23	307‡	126	108	*GENE P Blimp1	313‡	126	197	*mRNA E Otx	35‡	83	286	mRNA P Dpt	276‡	126	375	*PROTEIN E Brn	137	39‡
20	GENE E FoxO	310‡	126	109	*GENE P Bra	310‡	126	198	mRNA E Pks	162‡	68	287	*mRNA P Dri	121‡	80	376	PROTEIN E CAPK	298‡	126
21	GENE E FvMo	308‡	126	110	*GENE P Brn	307‡	126	199	*mRNA E Pmar1	59‡	77	288	mRNA P Endo16	97‡	77	377	*PROTEIN E cB	6‡	126
22	GENE E GataC	310‡	126	111	GENE P CAPK	318	126	200	mRNA E Sm27	263‡	126	289	mRNA P Erg	74‡	71	378	PROTEIN E CyP	297‡	126
23	*GENE E GataE	309‡	126	112	GENE P CyP	309‡	126	201	mRNA E Sm30	211‡	126	290	*mRNA P Ets1	23‡	81	379	*PROTEIN E Delta	131‡	126
24	GENE E Gcad	308‡	126	113	*GENE P Delta	310‡	126	202	mRNA E Sm50	238‡	126	291	*mRNA P Eve	132‡	64	380	*PROTEIN E Delta2	282‡	126
25	GENE E Gcm	312‡	126	114	GENE P Dpt	307‡	126	203	mRNA E Snail	275‡	126	292	mRNA P Ficolin	98‡	80	381	PROTEIN E Dpt	125	9‡
26	GENE E Gelsolin	307‡	126	115	*GENE P Dri	308‡	126	204	*mRNA E SoxB1	31‡	38	293	mRNA P FoxA	176‡	77	382	*PROTEIN E Dri	220‡	126
27	*GENE E HesC	308‡	126	116	GENE P Endo16	308‡	126	205	mRNA E SoxC	191‡	126	294	mRNA P FoxB	244	80‡	383	PROTEIN E Endo16	64‡	60
28	GENE E Hex	309‡	126	117	GENE P Erg	309‡	126	206	*mRNA E SuH	280	100‡	295	mRNA P FoxN23	228‡	126	384	PROTEIN E Erg	221‡	112
29	GENE E Hnf6	317‡	126	118	*GENE P Ets1	308‡	126	207	mRNA E SuTx	162‡	68	296	mRNA P FoxO	108‡	80	385	PROTEIN E ES	296‡	126
30	*GENE E Hox	311‡	126	119	*GENE P Eve	310‡	126	208	mRNA E TBR	266‡	116	297	mRNA P FvMo	271	68‡	386	*PROTEIN E Ets1	163‡	118
31	GENE E Kakapo	307‡	126	120	GENE P Ficolin	310‡	126	209	mRNA E Tel	189‡	126	298	mRNA P GataC	202	18‡	387	*PROTEIN E Eve	217	30‡
32	GENE E Lim	308‡	126	121	GENE P FoxA	313‡	126	210	mRNA E Tgif	254‡	115	299	*mRNA P GataE	47‡	78	388	PROTEIN E Ficolin	297‡	126
33	GENE E Msp130	313‡	126	122	GENE P FoxB	311‡	126	211	*mRNA E UbiqSoxB1	154	5‡	300	mRNA P Gcad	25	5‡	389	PROTEIN E FoxA	174‡	114
34	GENE E MspL	311‡	126	123	GENE P FoxN23	307‡	126	212	mRNA E UMR	281	102‡	301	mRNA P Gcm	161‡	117	390	PROTEIN E FoxB	279‡	126
35	GENE E Not	307‡	126	124	GENE P FoxO	310‡	126	213	*mRNA E UVAOtx	281	97‡	302	mRNA P Gelsolin	109‡	67	391	PROTEIN E FoxN23	287‡	126
36	GENE E Nrl	312‡	126	125	GENE P FvMo	308‡	126	214	mRNA E VEGF	283	96‡	303	*mRNA P HesC	155‡	72	392	PROTEIN E FoxO	297‡	126
37	GENE E OrCt	308‡	126	126	GENE P GataC	310‡	126	215	mRNA E VEGFR	248‡	126	304	mRNA P Hex	87‡	76	393	PROTEIN E frizzled a	318	126
38	*GENE E Otx	310‡	126	127	*GENE P GataE	309‡	126	216	*mRNA E Wnt8	43‡	61	305	mRNA P Hnf6	120‡	77	394	PROTEIN E frizzled i	318	52‡
39	GENE E Pks	308‡	126	128	GENE P Gcad	308‡	126	217	mRNA E z13	209‡	126	306	*mRNA P Hox	70‡	77	395	PROTEIN E FvMo	177	25‡
40	*GENE E Pmar1	309‡	126	129	GENE P Gcm	312‡	126	218	mRNA M Alx1	190‡	126	307	mRNA P Kakapo	109‡	67	396	PROTEIN E GataC	265	31‡
41	GENE E Sm27	313‡	126	130	GENE P Gelsolin	307‡	126	219	mRNA M Apobec	210	22‡	308	mRNA P L1	281	99‡	397	*PROTEIN E GataE	46‡	94
42	GENE E Sm30	307‡	126	131	*GENE P HesC	308‡	126	220	*mRNA M Blimp1	198	77‡	309	mRNA P Lim	201	77‡	398	PROTEIN E Gcad	68‡	126
43	GENE E Sm50	314‡	126	132	GENE P Hex	309‡	126	221	*mRNA M Bra	100‡	77	310	mRNA P Msp130	80‡	80	399	PROTEIN E Gcm	10‡	48
44	GENE E Snail	307‡	126	133	GENE P Hnf6	307‡	126	222	*mRNA M Brn	251	68‡	311	mRNA P MspL	63‡	80	400	PROTEIN E Gelsolin	75	15‡
45	*GENE E SoxB1	308‡	126	134	*GENE P Hox	311‡	126	223	mRNA M CAPK	107‡	67	312	mRNA P Not	274	68‡	401	*PROTEIN E Gro	300‡	126
46	GENE E SoxC	309‡	126	135	GENE P Kakapo	307‡	126	224	*mRNA M cB	152	5‡	313	mRNA P Nrl	105‡	66	402	*PROTEIN E GroTCF	143	45‡
47	GENE E SuTx	308‡	126	136	GENE P Lim	308‡	126	225	mRNA M CyP	270‡	126	314	mRNA P OrCt	226	62‡	403	PROTEIN E GroTFC	313	118‡
48	GENE E TBR	309‡	126	137	GENE P Msp130	313‡	126	226	*mRNA M Delta	50‡	47	315	*mRNA P Otx	37‡	82	404	PROTEIN E GSK3 a	318	85‡
49	GENE E Tel	309‡	126	138	GENE P MspL	311‡	126	227	mRNA M Dpt	85‡	67	316	mRNA P Pks	271	68‡	405	PROTEIN E GSK3 i	318	126
50	GENE E Tgif	310‡	126	139	GENE P Not	307‡	126	228	*mRNA M Dri	272‡	126	317	*mRNA P Pmar1	55‡	75	406	*PROTEIN E HesC	134‡	126
51	GENE E VEGFR	310‡	126	140	GENE P Nrl	312‡	126	229	mRNA M Endo16	91‡	77	318	m RNA P Sm27	67‡	80	407	PROTEIN E Hex	170‡	112
52	*GENE E Wnt8	309‡	126	141	GENE P OrCt	308‡	126	230	mRNA M Erg	257‡	126	319	mRNA P Sm30	54‡	44	408	PROTEIN E Hnf6	241‡	126
53	GENE E z13	309‡	126	142	*GENE P Otx	310‡	126	231	*mRNA M Ets1	168‡	126	320	mRNA P Sm50	60‡	80	409	*PROTEIN E Hox	17‡	104
54	GENE M Alx1	311‡	126	143	GENE P Pks	308‡	126	232	*mRNA M Eve	126‡	65	321	mRNA P Snail	111‡	69	410	PROTEIN E Kakapo	75	15‡
55	GENE M Apobec	309‡	126	144	*GENE P Pmar1	309‡	126	233	mRNA M Ficolin	267‡	126	322	*mRNA P SoxB1	182‡	126	411	PROTEIN E L1	303‡	126
56	*GENE M Blimp1	313‡	126	145	GENE P Sm27	313‡	126	234	mRNA M FoxA	180‡	77	323	mRNA P SoxC	231	77‡	412	PROTEIN E Lim	199‡	126
57	*GENE M Bra	310‡	126	146	GENE P Sm30	307‡	126	235	mRNA M FoxB	268‡	126	324	mRNA P SuTx	271	68‡	413	PROTEIN E Msp130	297‡	126
58	*GENE M Brn	307‡	126	147	GENE P Sm50	314‡	126	236	mRNA M FoxN23	230‡	110	325	mRNA P TBR	208	73‡	414	PROTEIN E MspL	296‡	126
59	GENE M CAPK	307‡	126	148	GENE P Snail	307‡	126	237	mRNA M FoxO	267‡	126	326	mRNA P Tel	218	77‡	415	*PROTEIN E nBTCF	83‡	112
60	GENE M CyP	309‡	126	149	*GENE P SoxB1	308‡	126	238	mRNA M FvMo	165‡	68	327	mRNA P Tgif	95‡	74	416	PROTEIN E Not	253	25‡
61	*GENE M Delta	310‡	126	150	GENE P SoxC	309‡	126	239	mRNA M GataC	216	68‡	328	mRNA P UbiqAlx1	150	6‡	417	*PROTEIN E Notch	133‡	59
62	GENE M Dpt	308‡	126	151	GENE P SuTx	308‡	126	240	*mRNA M GataE	113‡	75	329	mRNA P UbiqES	154	6‡	418	*PROTEIN E Notch2	259	42‡
63	*GENE M Dri	308‡	126	152	GENE P TBR	309‡	126	241	mRNA M Gcad	25	5‡	330	*mRNA P UbiqEts1	154	4‡	419	PROTEIN E Nrl	242‡	126
64	GENE M Endo16	308‡	126	153	GENE P Tel	309‡	126	242	mRNA M Gcm	86‡	57	331	*mRNA P UbiqHesC	154	7‡	420	PROTEIN E OrCt	286	35‡
65	GENE M Erg	309‡	126	154	GENE P Tgif	310‡	126	243	mRNA M Gelsolin	102	20‡	332	mRNA P UbiqHnf6	154	5‡	421	*PROTEIN E Otx	14‡	98
66	*GENE M Ets1	308‡	126	155	GENE P VEGFR	310‡	126	244	*mRNA M HesC	173‡	105	333	mRNA P UbiqSoxC	152	5‡	422	PROTEIN E Pks	177	25‡
67	*GENE M Eve	310‡	126	156	*GENE P Wnt8	309‡	126	245	mRNA M Hex	250‡	126	334	mRNA P UbiqTel	152	5‡	423	*PROTEIN E Pmar1	21‡	126
68	GENE M Ficolin	310‡	126	157	GENE P z13	309‡	126	246	mRNA M Hnf6	193‡	126	335	mRNA P VEGFR	32‡	79	424	PROTEIN E Sm27	297‡	126
69	GENE M FoxA	313‡	126	158	mRNA E Alx1	215‡	119	247	*mRNA M Hox	72‡	77	336	*mRNA P Wnt8	45‡	61	425	PROTEIN E Sm30	295‡	126
70	GENE M FoxB	311‡	126	159	mRNA E Apobec	224	62‡	248	mRNA M Kakapo	102	20‡	337	mRNA P z13	203‡	126	426	PROTEIN E Sm50	295‡	126
71	GENE M FoxN23	307‡	126	160	*mRNA E Blimp1	196	74‡	249	mRNA M Lim	181‡	77	338	none	57‡	126	427	PROTEIN E Snail	166‡	126
72	GENE M FoxO	310‡	126	161	*mRNA E Bra	92‡	76	250	mRNA M Msp130	262‡	126	339	*PRE E cB	316‡	126	428	*PROTEIN E SoxB1	12‡	126
73	GENE M FvMo	308‡	126	162	*mRNA E Brn	245	95‡	251	mRNA M MspL	225‡	126	340	PRE E Gcad	315‡	126	429	PROTEIN E SoxC	246‡	126
74	GENE M GataC	310‡	126	163	mRNA E CAPK	307‡	126	252	mRNA M Not	252	68‡	341	*PRE E Notch	316‡	126	430	*PROTEIN E SuH	142‡	89
75	*GENE M GataE	309‡	126	164	*mRNA E cB	157	5‡	253	*mRNA M Notch	154	5‡	342	*PRE E Otx	315‡	126	431	*PROTEIN E SuHN	103	10‡
76	GENE M Gcad	308‡	126	165	mRNA E CyP	270‡	126	254	mRNA M Nrl	156	3‡	343	*PRE E SoxB1	315‡	126	432	PROTEIN E SuTx	177	25‡
77	GENE M Gcm	312‡	126	166	*mRNA E Delta	164‡	117	255	mRNA M OrCt	210	22‡	344	*PRE E SuH	316‡	126	433	PROTEIN E TBR	206‡	119
78	GENE M Gelsolin	308‡	126	167	mRNA E Dpt	77‡	40	256	*mRNA M Otx	34‡	80	345	*PRE E UbiqSoxB1	316‡	126	434	*PROTEIN E TCF	139	53‡
79	*GENE M HesC	308‡	126	168	*mRNA E Dri	266‡	126	257	mRNA M Pks	153	21‡	346	PRE E UMR	316‡	126	435	PROTEIN E Tel	235‡	126
80	GENE M Hex	309‡	126	169	mRNA E Endo16	90‡	58	258	*mRNA M Pmar1	56‡	77	347	*PRE E UVAOtx	316‡	126	436	PROTEIN E Tgif	204‡	111
81	GENE M Hnf6	317‡	126	170	mRNA E Erg	261‡	109	259	mRNA M Sm27	260‡	126	348	PRE E VEGF	316‡	126	437	PROTEIN E UbiqAlx1	309	110‡
82	*GENE M Hox	311‡	126	171	mRNA E ES	185‡	126	260	mRNA M Sm30	211‡	126	349	*PRE M cB	316‡	126	438	*PROTEIN E UbiqDelta	310	114‡
83	GENE M Kakapo	308‡	126	172	*mRNA E Ets1	171‡	116	261	mRNA M Sm50	232‡	126	350	PRE M Gcad	315‡	126	439	PROTEIN E UbiqES	309‡	126
84	GENE M Lim	308‡	126	173	*mRNA E Eve	147‡	63	262	mRNA M Snail	275‡	126	351	*PRE M Notch	316‡	126	440	*PROTEIN E UbiqEts1	308	108‡
85	GENE M Msp130	313‡	126	174	mRNA E Ficolin	267‡	126	263	*mRNA M SoxB1	31‡	38	352	*PRE M Otx	315‡	126	441	PROTEIN E UbiqGcad	308	106‡
86	GENE M MspL	311‡	126	175	mRNA E FoxA	175‡	78	264	mRNA M SoxC	191‡	126	353	*PRE M SoxB1	315‡	126	442	*PROTEIN E UbiqHesC	308‡	126
87	GENE M Not	307‡	126	176	mRNA E FoxB	268‡	126	265	*mRNA M SuH	281	102‡	354	*PRE M SuH	316‡	126	443	PROTEIN E UbiqHnf6	317‡	126
88	GENE M Nrl	313‡	126	177	mRNA E FoxN23	234‡	126	266	mRNA M SuTx	165‡	68	355	*PRE M UbiqSoxB1	316‡	126	444	*PROTEIN E UbiqSoxB1	27‡	126
89	GENE M OrCt	309‡	126	178	mRNA E FoxO	267‡	126	267	mRNA M TBR	269‡	110	356	*PRE M UMADelta	316‡	126	445	PROTEIN E UbiqSoxC	309‡	126

Table 1: Node names and associated identification numbers (IDs) (assigned in alphabetical order) for the endomesoderm network. Table is read from top to bottom and from left to right. Ψ_P and Ψ_L are the ranks of PANI and LSA, respectively. The different embryonic region are represented by M, E and P which indicates mesoderm, endoderm and PMC cells, respectively. Nodes associated to molecules found in Figure 1a are marked with *. The higher normalized ranks of each node i is marked with ‡, where the normalized PANI and LSA ranks are $\frac{\Psi_{P_i}}{MAX(\Psi_P)}$ and $\frac{\Psi_{L_i}}{MAX(\Psi_L)}$, respectively; Ψ_{P_i} is the PANI rank of node i and $MAX(\Psi_P)$ is the maximum PANI rank of all nodes in Tables 1 and 2. This table contains IDs [1 – 445] and the rest of the IDs continue in Table 2.

ID	Node Name	Ψ_P	Ψ_L	ID	Node Name	Ψ_P	Ψ_L	ID	Node Name	Ψ_P	Ψ_L	ID	Node Name	Ψ_P	Ψ_L
446	protein E UbiqTel	309‡	126	491	*protein M HesC	159‡	114	536	protein M vegfr	140‡	126	581	protein P Msp130	101‡	126
447	*protein E umADelta	310	114‡	492	protein M Hex	172‡	117	537	protein M vegfSignal	124‡	114	582	protein P MspL	89‡	126
448	protein E umANrl	312‡	126	493	protein M Hnf6	243‡	126	538	*protein M Wnt8	2‡	21	583	*protein P nBtcf	76‡	113
449	protein E umr	304	90‡	494	*protein M Hox	22‡	110	539	protein M z13	293‡	126	584	protein P Not	277	26‡
450	*protein E uvaOtx	291	93‡	495	protein M Kakapo	71	17‡	540	protein P Alx1	122‡	108	585	*protein P Notch	151‡	126
451	protein E vegf	82‡	86	496	protein M L1	303‡	126	541	protein P Apobec	288	34‡	586	*protein P Notch2	119‡	126
452	protein E vegfr	141‡	126	497	protein M Lim	200‡	126	542	*protein P Blimp1	136‡	126	587	protein P Nrl	179	8‡
453	protein E vegfSignal	124‡	114	498	protein M Msp130	297‡	126	543	*protein P Bra	24‡	126	588	protein P OrCt	288	34‡
454	*protein E Wnt8	1‡	19	499	protein M MspL	295‡	126	544	*protein P Brn	184	26‡	589	*protein P Otx	16‡	123
455	protein E z13	294‡	126	500	*protein M nBtcf	78‡	126	545	protein P capk	298‡	126	590	protein P Pks	277	26‡
456	protein gcm	310‡	126	501	protein M Not	255	27‡	546	*protein P cB	7‡	126	591	*protein P Pmar1	41‡	110
457	protein M Alx1	236‡	126	502	*protein M Notch	49‡	125	547	protein P CyP	106‡	126	592	protein P Sm27	104‡	126
458	protein M Apobec	285	36‡	503	*protein M Notch2	197	43‡	548	*protein P Delta	48‡	126	593	protein P Sm30	66	2‡
459	*protein M Blimp1	123‡	126	504	protein M Nrl	188	41‡	549	*protein P Delta2	79‡	50	594	protein P Sm50	96‡	126
460	*protein M Bra	38‡	126	505	protein M OrCt	285	36‡	550	protein P Dpt	297‡	126	595	protein P Snail	9‡	11
461	*protein M Brn	145	27‡	506	*protein M Otx	13‡	122	551	*protein P Dri	36‡	110	596	*protein P SoxB1	160‡	126
462	protein M capk	75	15‡	507	protein M Pks	129	23‡	552	protein P Endo16	84‡	126	597	protein P SoxC	186‡	126
463	*protein M cB	5‡	126	508	*protein M Pmar1	20‡	126	553	protein P Erg	33‡	114	598	*protein P SuH	299‡	126
464	protein M CyP	297‡	126	509	protein M Sm27	297‡	126	554	*protein P Ets1	18‡	120	599	*protein P SuHN	128‡	109
465	*protein M Delta	39	12‡	510	protein M Sm30	295‡	126	555	*protein P Eve	222	32‡	600	protein P SuTx	277	26‡
466	*protein M Delta2	229	1‡	511	protein M Sm50	295‡	126	556	protein P Ficolin	88‡	126	601	protein P Tbr	99‡	126
467	protein M Dpt	112	16‡	512	protein M Snail	166‡	126	557	protein P FoxA	183‡	126	602	*protein P tcf	138‡	55
468	*protein M Dri	239‡	126	513	*protein M SoxB1	11‡	126	558	protein P FoxB	219‡	126	603	protein P Tel	149‡	126
469	protein M Endo16	65‡	126	514	protein M SoxC	246‡	126	559	protein P FoxN23	289‡	126	604	protein P Tgif	26‡	107
470	protein M Erg	223‡	117	515	*protein M SuH	130‡	88	560	protein P FoxO	94‡	126	605	protein P UbiqAlx1	62‡	125
471	*protein M Ets1	169‡	112	516	*protein M SuHN	93‡	56	561	protein P frizzled a	318	126	606	*protein P UbiqDelta	310‡	126
472	*protein M Eve	213	28‡	517	protein M SuTx	194	33‡	562	protein P frizzled i	318	52‡	607	protein P UbiqEs	81‡	126
473	protein M Ficolin	297‡	126	518	protein M Tbr	212‡	115	563	protein P FvMo	277	26‡	608	*protein P UbiqEts1	28‡	124
474	protein M FoxA	192‡	126	519	*protein M tcf	135‡	54	564	protein P GataC	256	24‡	609	protein P UbiqGcad	308	113‡
475	protein M FoxB	278‡	126	520	protein M Tel	233‡	126	565	*protein P GataE	44‡	126	610	*protein P UbiqHesC	30‡	121
476	protein M FoxN23	290‡	115	521	protein M Tgif	240‡	107	566	protein P Gcad	68‡	126	611	protein P UbiqHnf6	8‡	126
477	protein M FoxO	297‡	126	522	protein M UbiqAlx1	311‡	126	567	protein P Gcm	110‡	114	612	*protein P UbiqSoxB1	308‡	126
478	protein M frizzled a	318	126	523	*protein M UbiqDelta	310‡	126	568	protein P Gelsolin	69	14‡	613	protein P UbiqSoxC	52‡	126
479	protein M frizzled i	318	52‡	524	protein M UbiqEs	318	126	569	*protein P Gro	302‡	126	614	protein P UbiqTel	52‡	126
480	protein M FvMo	194	33‡	525	*protein M UbiqEts1	308‡	126	570	*protein P Grotcf	148	46‡	615	*protein P umADelta	310‡	126
481	protein M GataC	264	29‡	526	protein M UbiqGcad	308	106‡	571	protein P Grotfc	313‡	126	616	protein P umANrl	312‡	126
482	*protein M GataE	51‡	111	527	*protein M UbiqHesC	308	117‡	572	protein P gsk3 a	318	85‡	617	*protein P uvaOtx	310	119‡
483	protein M Gcad	68‡	126	528	protein M UbiqHnf6	317‡	126	573	protein P gsk3 i	318	126	618	protein P vegfr	116‡	119
484	protein M Gcm	29‡	49	529	*protein M UbiqSoxB1	27‡	126	574	*protein P HesC	118‡	114	619	protein P vegfSignal	53	13‡
485	protein M Gelsolin	71	17‡	530	protein M UbiqSoxC	309‡	126	575	protein P Hex	15‡	116	620	*protein P Wnt8	3‡	21
486	*protein M Gro	301‡	126	531	protein M UbiqTel	309‡	126	576	protein P Hnf6	40‡	126	621	protein P z13	293‡	126
487	*protein M Grotcf	144	46‡	532	*protein M umADelta	284	91‡	577	*protein P Hox	19‡	126	622	ribosome	318	126
488	protein M Grotfc	313‡	126	533	protein M umANrl	292	92‡	578	protein P Kakapo	69	14‡				
489	protein M gsk3 a	318	85‡	534	protein M umr	304	90‡	579	protein P L1	158‡	87				
490	protein M gsk3 i	318	126	535	*protein M uvaOtx	310‡	126	580	protein P Lim	207‡	126				

Table 2: Continuation of node names and associated IDs for the endomesoderm network from Table 1. Table is read from top to bottom and from left to right. Explanations of symbols follow that in Table 1.

We perform ROC analysis to examine the enrichment of the benchmark `Endo16` regulators in the top k nodes prioritized by PANI, PSSD, TDE and BC (properties used to compute the putative target scores). We vary k in the range $[0 - |V|]$, where $|V|$ is the size of the endomesoderm network. The area under the ROC curve (AUC) (Figure 2) is 0.625, 0.56, 0.572 and 0.637 for PANI, PSSD, TDE and BC, respectively. In the case of the endomesoderm network, the performance of PANI is mainly attributed to BC. This is probably because unlike PSSD, BC is less sensitive to experimental error and parameter estimation. Also, the unique topological characteristics of the network as discussed earlier contribute to the important role BC plays in this network. In summary, the good performances of PANI and BC indicate that network topology features are useful complement to traditional simulation-based model analysis, especially for networks where the dynamics are still fuzzy. Note that PANI achieves slight improvement over BC in terms of the minimum number of top scoring nodes required to identify the benchmark regulators (*MinNode*) (PANI=599, BC=603) and the enrichment of benchmark regulatory nodes in the top-61 ranked nodes (PANI=74%, BC=65.6%).

4.2 Comparison with Random Prioritization and Local Sensitivity Analysis (LSA)

For random prioritization, the set of nodes related to the benchmark `Endo16` regulators are randomly prioritized 100 times. For simplicity, we assume that the random prioritization assigns a unique rank from the range [1 – 622] to each benchmark node. We compare the minimum number of top scoring nodes required to identify all the benchmark regulators (*MinNode*). The *MinNode* of PANI is 599 while that of the random trials varies in the range [612 – 622]. Hence, PANI can identify the benchmark `Endo16` regulators using much fewer top scoring nodes compared to random prioritization. Next, we perform a paired t-test of ranked nodes generated by PANI and random prioritization. The rankings are normalized to the range [0 – 1] before carrying out the paired t-test to account for the presence of ties in PANI's rankings and the lack of ties in the random rankings. The p-value of the two-tailed paired t-test varies in the range $[1.54\times10^{-5} - 0.1]$, suggesting that the rankings of PANI and the random trials are different. Furthermore, PANI ranks benchmark regulators higher than random trials at 5% significance level in one-tailed paired t-test and the ROC AUC varies in the range [0.44 – 0.54].

We use *Copasi* to perform the LSA and set the parameters as follows: {*Subtask=Time Series*, *Function=Non-Constant Concentrations of Species*, *Variable=Initial Concentrations*}. The analysis took ∼19 minutes and the rankings are represented in Tables 1 and 2. The Spearman's correlation coefficient between PANI's and LSA's ranks is 0.472, implying a moderate correlation between the rankings. The *MinNode* values of PANI and LSA are 599 and 622, respectively, implying that PANI requires fewer top ranking nodes to identify all the benchmark regulators. The one-tailed paired t-test performed on the normalized rankings of PANI and LSA reveals that PANI ranks benchmark regulators higher than LSA at 5% significance level. In fact, PANI ranks 80.1% of the benchmark regulators higher than LSA. For instance, compared to PANI, LSA ranks all nodes associated to `Wnt8` and `Bra` lower although both are `Endo16` regulators. Further-

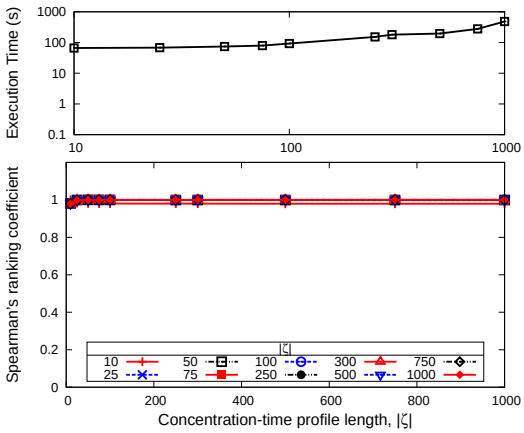

Figure 3: Effect of $|\zeta|$ on runtime and rankings.

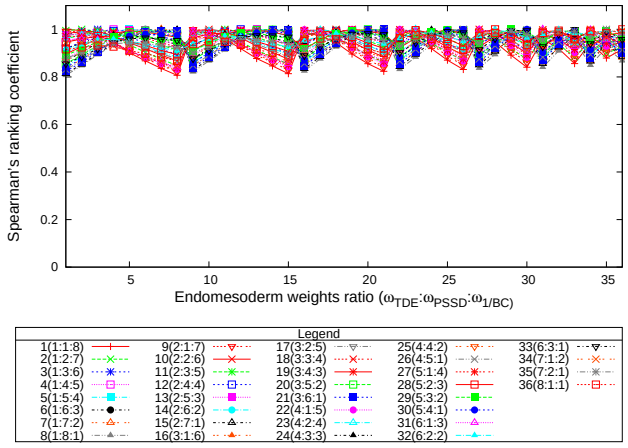

Figure 4: Effect of property-weights on relative ranking result of nodes.

more, the ROC AUC of LSA is 0.549 (Figure 2) and in the LSA's top-61 nodes, only 20 (32.8%) are in the set of benchmark nodes. Hence, PANI produces superior prioritization results compared to random prioritization and LSA.

5. ROBUSTNESS OF PRIORITIZATION

In this section, we study the robustness of the target prioritization step (Step 1 in Section 4) by examining the effect of various parameters. The parameters that we examine are the concentration-time profile length ($|\zeta|$), the weights of the three properties (ω_{PSSD}, ω_{TDE} and ω_{BC}), and the node of interest (output node). Recall that the concentration-time profile is used to compute PSSD while the weights are used for the calculation of the putative target score. The output node is used as a reference for the reachability-based pruning of non-regulators and the computation of PSSD. We vary each of these parameters and examine their effects on the prioritization ranking as well as the execution time of Step 1. Note that examining the effects of the parameters on ranking allows us to study the sensitivity of the prioritization results to these parameters, giving us a sense of the robustness of PANI-based targets prioritization in the endomesoderm network.

5.1 Effects of Profile Length ($|\zeta|$)

In this experiment, we examine the effect of varying the number of time points in the concentration-time profile ζ

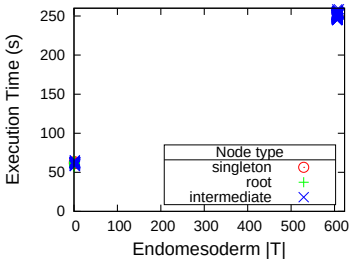

Figure 5: Relationship of execution time (s) and $|T|$.

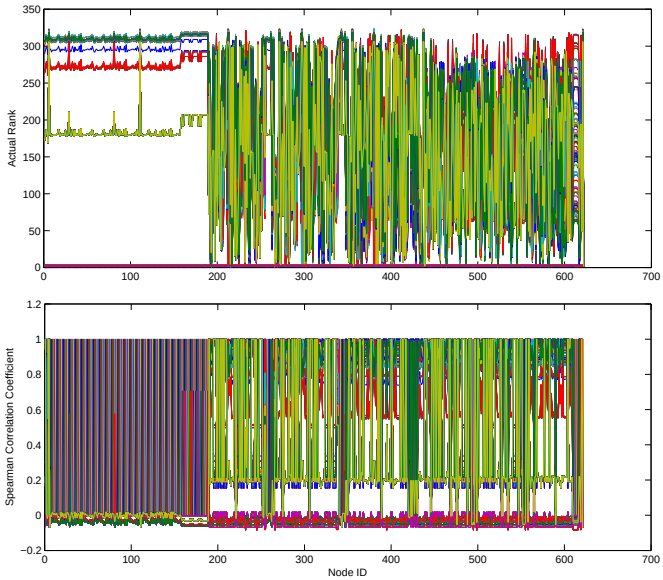

Figure 6: Effect of varying output node on endomesoderm ranking results. Node names of the corresponding node ID can be found in Tables 1 and 2.

($|\zeta|$) on the ranking. The profiles are obtained using *Copasi* where $|\zeta|$ varies in the range of {10, 25, 50, 75, 100, 250, 300, 500, 750, 1000}. We observe that the execution times of PANI increase with increasing value of $|\zeta|$ (Figure 3) as the latter affects the time for calculating PSSD.

Next, we investigate the effect of $|\zeta|$ on the ranking results. This gives us a sense of the minimum $|\zeta|$ required to produce superior quality ranking and allows us to assess the practical execution time more accurately. We compare the changes in the ranking results using Spearman's ranking correlation coefficient as depicted in Figure 3. We observe that $|\zeta| = 10$ has a lower coefficient with respect to the rankings obtained for other values of $|\zeta|$. Although the coefficient at $|\zeta| = 10$ is lower, it is still relatively high at ∼98%, suggesting that the concentration-time profiles in the endomesoderm network may have few profile changes and a small $|\zeta|$ is sufficient to capture the variations in the profiles. In fact, three of the benchmark regulators {`PROTEIN E Pmar1`, `PROTEIN M Hox`, `PROTEIN M Pmar1`} are assigned the same ranks and the standard deviation of the ranks of the benchmark regulators vary in the range [0 – 8] across the entire range of $|\zeta|$. At $|\zeta| > 25$, the correlation coefficient approaches a constant value of ∼ 100% when compared with other values of $|\zeta| > 25$. Hence, a small value of $|\zeta|$ is sufficient and the execution time of Step 1 for $|\zeta| < 100$ is less than 100 seconds.

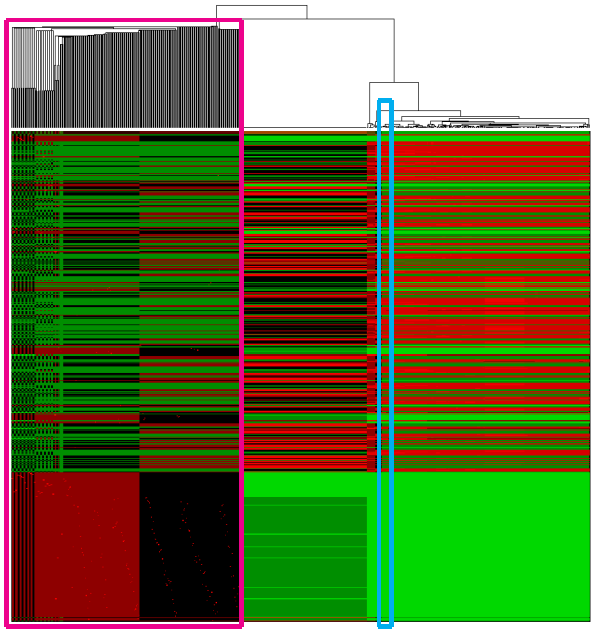

Figure 7: Clustergram analysis of Spearman's correlation coefficient of endomesoderm ranking result when output node is varied.

5.2 Effects of Weights (ω_{PSSD}, ω_{TDE} and $\omega_{\frac{1}{BC}}$)

We now investigate the effects of different scalar weight factors on the ranking result by examining how the percentage of common putative target nodes varies as the weights are modified. We vary each weight in the range of $\{0.1, 0.2, 0.3, 0.4, 0.5, 0.6, 0.7, 0.8, 0.9\}$ while ensuring that $\omega_{\text{PSSD}} + \omega_{\text{TDE}} + \omega_{\frac{1}{\text{BC}}} = 1$. This produces 36 different weight-ratios. For each weight-ratio, the putative target score of each node is calculated. Then, the Spearman's correlation coefficient of the rankings of each pair of weight-ratios is evaluated. We find that many of the weight-ratios contain common putative target nodes. The correlation coefficient ranges between $\sim$ 0.8 to 1 (Figure 4) and 76.9% of the top-50% putative target nodes in all the ratios are common. Next, we look at the minimum number of top scoring nodes required to identify the benchmark `Endo16` regulators (*MinNode*) in these weight-ratios. We find that the *MinNode* needed to identify at least 75% of the benchmark regulators for this 622-node network is 497. The standard deviation of the ranks of the benchmark regulators vary in the range [0.58 – 84.7] across the entire range of weight-ratios and 63.3% of the regulators has deviation of less than 20. In particular, `PROTEIN M SoxB1`, `PROTEIN E SoxB1`, `PROTEIN E cB`, `PROTEIN M cB`, `PROTEIN P cB`, `PROTEIN E Hox` and `PROTEIN P Hox` are consistently ranked in the 90^{th} percentile. These results imply that although the rankings of the targets vary, most of the targets are still ranked high enough to be considered as a putative target node in most weight-ratios, and many of these putative target nodes correspond to the benchmark regulators.

5.3 Effects of Selecting Different Output Node

In this experiment, we examine the effect of selecting a different output node on the execution time and the prioritization results. When we vary the output node, the number of pruned targets $|T|$ obtained from the pruning phase (Section 3.1) falls into two distinct clusters (Figure 5), one containing less than 20 nodes (cluster 1) and another containing more than 600 nodes (cluster 2). This distribution of $|T|$ is likely due to the network structure such as the presence of SCCs (Section 2.1). Recall that the endomesoderm network contains a large SCC with 360 nodes. Since nodes in the same SCC have the same set of pruned targets and hence the same $|T|$, it is likely that selecting an output node belonging to this SCC contributes to many of the points in cluster 2. Observe that the execution time varies linearly with $|T|$. When we vary the output node, the prioritization results change. We perform Spearman's rank correlation coefficient and clustergram analysis to investigate the effect of selecting different output node on the prioritization results. For the purpose of computing the Spearman's ranked correlation coefficient, candidate nodes that are pruned ($V \backslash T$) are assigned the lowest rank value to reflect their low relevance as putative target node.

Although the endomesoderm network (Figure 6) appears to have a close Spearman's correlation coefficient across the entire range of output nodes, some of these output nodes seem to share more similar rank correlation coefficient than others. The endomesoderm network's correlation coefficient appears to fall into two different clusters. The clustergram analysis (Figure 7) reveals two main clusters. The first main cluster (Figure 7, magenta box) contains the set of root nodes, singleton nodes and intermediate nodes which are not in any SCC; the second main cluster contains nodes in the 8 two-node SCCs and the 360-node SCC. For the two-node SCCs, nodes in the same SCC were clustered together. For the larger-sized SCC, nodes of the same types tend to form sub-clusters. For instance, nodes associated to `Blimp1` and `FoxA` cluster together to form a sub-cluster {`mRNA E Blimp1`, `mRNA E FoxA`, `mRNA M Blimp1`, `mRNA M FoxA`, `mRNA P Blimp1`, `mRNA P FoxA`, `mRNA P GataC`} (Figure 7, blue box). Hence, for the endomesoderm network, selection of output nodes in the same SCC produces closer rank correlation coefficient and hence more similar prioritization results. This is most likely due to output nodes in the same SCC sharing similar PSSD as time series profiles of genes in the same module are highly correlated in gene regulatory network [35].

6. CONCLUSIONS

In this paper, we apply prioritization tools (LSA and PANI) to the sea urchin endomesoderm gene regulatory network to identify putative target nodes involved in the regulation of `Endo16`. Prioritization tools assist researchers in identifying a set of nodes that should be prioritized for the study of a particular problem, thus saving precious time and resources. Target prioritization is particularly useful for large networks where visualization is challenging and manually analyzing the network is virtually impossible. We obtain a prioritized list of nodes that corresponds well with the set of benchmark `Endo16` regulators using PANI in around 250 seconds. We find that the characteristics of the endomesoderm network affect PANI's performance. Specifically, the presence of a large SCC and constant concentration profiles of many nodes significantly reduced the roles played by TDE and PSSD features for identifying target molecules. This highlights an intricate relationship between the network characteristics and its influence on the role of structural and dynamic properties of nodes in *in silico* targets prioritization, which should be considered in future applications.

Besides identifying the benchmark `Endo16` regulators, PANI also prioritizes several nodes (*e.g.*, `Snail`) that play a regulatory role for `Endo16` but are not in the set of benchmark nodes. Hence, we can exploit the capability of *in silico* target prioritization techniques (*e.g.*, PANI) to identify these interesting nodes to gain further biological insights, such as improving on the `Endo16` regulatory pathway which is far from complete. For instance, we can design experiments to uncover the relationships between nodes that PANI prioritizes and the `Endo16` benchmark regulators to help us fill the gaps in the pathway, thereby improving its accuracy.

7. ACKNOWLEDGMENTS

The authors are supported by grant from the Singapore-MIT Alliance Programme in Computational and Systems Biology.

8. REFERENCES

[1] G. Amore et al. Spdeadringer, a sea urchin embryo gene required separately in skeletogenic and oral ectoderm gene regulatory networks. *Developmental Biology*, 261(1):55 – 81, 2003.

[2] L. Angerer et al. Mutual antagonism of soxb1 and canonical wnt signaling in sea urchin embryos. *Signal Transduction*, 7:174–180, 2007.

[3] C. Arenas-Mena et al. Hindgut specification and cell-adhesion functions of sphox11/13b in the endoderm of the sea urchin embryo. *Development, Growth & Differentiation*, 48(7):463–472, 2006.

[4] S. Ben-Tabou de Leon et al. Deciphering the underlying mechanism of specification and differentiation: The sea urchin gene regulatory network. *Sci. STKE*, 2006(361):pe47, 2006.

[5] U. Brandes. A faster algorithm for betweenness centrality. *Journal of Mathematical Sociology*, 25:163–177, 2001.

[6] C. Byrum et al. Blocking dishevelled signaling in the noncanonical wnt pathway in sea urchins disrupts endoderm formation and spiculogenesis, but not secondary mesoderm formation. *Developmental Dynamics*, 238(7):1649–1665, 2009.

[7] H. Chua et al. Pani: A novel algorithm for fast discovery of putative target nodes in signaling networks. In *ACM Conference on Bioinformatics, Computational Biology and Biomedicine*, 2011.

[8] E. Davidson et al. A Genomic Regulatory Network for Development. *Science*, 295(5560):1669–1678, 2002.

[9] C. Ettensohn. Gastrulation in the sea urchin embryo is accompanied by the rearrangement of invaginating epithelial cells. *Developmental Biology*, 112(2):383 – 390, 1985.

[10] V. Hinman et al. Developmental gene regulatory network architecture across 500 million years of echinoderm evolution. *Proceedings of the National Academy of Sciences*, 100(23):13356–13361, 2003.

[11] V. Hinman et al. Evolutionary plasticity of developmental gene regulatory network architecture. *Proceedings of the National Academy of Sciences*, 104(49):19404–19409, 2007.

[12] D. Hu et al. Time-dependent sensitivity analysis of biological networks: Coupled MAPK and PI3K signal transduction pathways. *The Journal of Physical Chemistry A*, 110(16):5361–5370, 2006.

[13] W.-C. Hwang et al. Identification of information flow-modulating drug targets: a novel bridging paradigm for drug discovery. *Clin Pharmacol Ther*, 84(5):563–572, Nov 2008.

[14] C. Jopling et al. Shp2 knockdown and noonan/leopard mutant shp2Ŭinduced gastrulation defects. *PLoS Genet*, 3(12):e225, 12 2007.

[15] E. Keogh et al. Derivative dynamic time warping. In *In First SIAM International Conference on Data Mining (SDMŠ2001*, 2001.

[16] C. Kuhn et al. Monte carlo analysis of an ode model of the sea urchin endomesoderm network. *BMC Systems Biology*, 3(1):83, 2009.

[17] E. Lai et al. Drosophila neuralized is a ubiquitin ligase that promotes the internalization and degradation of delta. *Developmental Cell*, 1(6):783 – 794, 2001.

[18] N. Le Novère et al. Biomodels database: a free, centralized database of curated, published, quantitative kinetic models of biochemical and cellular systems. *Nucleic Acids Res*, 34(Database issue):D689–D691, Jan 2006.

[19] T. Lee et al. Transcriptional regulatory networks in saccharomyces cerevisiae. *Science*, 298(5594):799–804, 2002.

[20] C. Livi et al. Expression and function of blimp1/krox, an alternatively transcribed regulatory gene of the sea urchin endomesoderm network. *Developmental Biology*, 293(2):513 – 525, 2006.

[21] C. Logan et al. Nuclear beta-catenin is required to specify vegetal cell fates in the sea urchin embryo. *Development*, 126(2):345–357, 1999.

[22] S. Materna et al. Logic of gene regulatory networks. *Curr Opin Biotechnol*, 18(4):351–354, Aug 2007.

[23] C. Nocente-McGrath et al. Endo16, a lineage-specific protein of the sea urchin embryo, is first expressed just prior to gastrulation. *Developmental Biology*, 136(1):264 – 272, 1989.

[24] P. Oliveri et al. A regulatory gene network that directs micromere specification in the sea urchin embryo. *Developmental Biology*, 246(1):209 – 228, 2002.

[25] P. Oliveri et al. Global regulatory logic for specification of an embryonic cell lineage. *Proceedings of the National Academy of Sciences*, 105(16):5955–5962, 2008.

[26] O. Otim et al. Sphnf6, a transcription factor that executes multiple functions in sea urchin embryogenesis. *Developmental Biology*, 273(2):226 – 243, 2004.

[27] D. Pant et al. Automated oncogene detection in complex protein networks with applications to the mapk signal transduction pathway. *Biophysical Chemistry*, 113(3):275 – 288, 2005.

[28] I. Peter et al. The endoderm gene regulatory network in sea urchin embryos up to mid-blastula stage. *Developmental Biology*, 340(2):188 – 199, 2010. Special Section: Gene Regulatory Networks for Development.

[29] C. Roberts et al. Targeted mutagenesis of the hira gene results in gastrulation defects and patterning abnormalities of mesoendodermal derivatives prior to early embryonic lethality. *Mol. Cell. Biol.*, 22(7):2318–2328, 2002.

[30] L. Romano et al. Endo16 is required for gastrulation in the sea urchin lytechinus variegatus. *Dev Growth Differ*, 48(8):487–497, Oct 2006.

[31] E. Röttinger et al. A Raf/MEK/ERK signaling pathway is required for development of the sea urchin embryo micromere lineage through phosphorylation of the transcription factor Ets. *Development*, 131(5):1075–1087, 2004.

[32] J. Smith et al. A spatially dynamic cohort of regulatory genes in the endomesodermal gene network of the sea urchin embryo. *Developmental Biology*, 313(2):863 – 875, 2008.

[33] I. Sobolá. Global sensitivity indices for nonlinear mathematical models and their Monte Carlo estimates. *Math. Comput. Simul.*, 55(1-3):271–280, 2001.

[34] L. Solnica-Krezel. Conserved patterns of cell movements during vertebrate gastrulation. *Curr Biol*, 15(6):R213–R228, Mar 2005.

[35] S. Tornow et al. Functional modules by relating protein interaction networks and gene expression. *Nucleic Acids Res*, 31(21):6283–6289, Nov 2003.

[36] L. Wolpert. Gastrulation and the evolution of development. *Development*, 116(Supplement):7–13, Apr 1992.

[37] S.-Y. Wu et al. The snail repressor is required for pmc ingression in the sea urchin embryo. *Development*, 134(6):1061–1070, 2007.

[38] C.-H. Yuh et al. Correct expression of spec2a in the sea urchin embryo requires both otx and other cis-regulatory elements. *Developmental Biology*, 232(2):424 – 438, 2001.

[39] C.-H. Yuh et al. An otx cis-regulatory module: a key node in the sea urchin endomesoderm gene regulatory network. *Developmental Biology*, 269(2):536 – 551, 2004.

[40] C.-H. Yuh et al. Brn1/2/4, the predicted midgut regulator of the endo16 gene of the sea urchin embryo. *Developmental Biology*, 281(2):286 – 298, 2005.

[41] Z. Zi et al. In silico identification of the key components and steps in IFN-gamma induced JAK-STAT signaling pathway. *FEBS Lett*, 579(5):1101–1108, Feb 2005.

[42] Z. Zi et al. A quantitative study of the hog1 mapk response to fluctuating osmotic stress in saccharomyces cerevisiae. *PLoS ONE*, 5(3):e9522, 03 2010.