Bioinformatics, 33(10), 2017, 1528–1535
doi: 10.1093/bioinformatics/btw796
Advance Access Publication Date: 29 December 2016
Original Paper

OXFORD

Systems biology

Inference of cellular level signaling networks using single-cell gene expression data in *Caenorhabditis elegans* reveals mechanisms of cell fate specification

Xiao-Tai Huang,[1,2,†] Yuan Zhu,[3,4,*,†] Lai Hang Leanne Chan,[2] Zhongying Zhao[5] and Hong Yan[2]

[1]School of Computer Science and Technology, Xidian University, Xi'an, Shaanxi, China, [2]Department of Electronic Engineering, City University of Hong Kong, Hong Kong, China, [3]School of Automation, China University of Geosciences, Wuhan, China, [4]Hubei Key Laboratory of Advanced Control and Intelligent Automation for Complex Systems and [5]Department of Biology, Hong Kong Baptist University, Hong Kong, China

*To whom correspondence should be addressed.
[†]The authors wish it to be known that, in their opinion, the first two authors should be regarded as joint First Authors.
Associate Editor: Cenk Sahinalp

Received on June 20, 2016; revised on November 18, 2016; editorial decision on December 12, 2016; accepted on December 14, 2016

Abstract

Motivation: Cell fate specification plays a key role to generate distinct cell types during metazoan development. However, most of the underlying signaling networks at cellular level are not well understood. Availability of time lapse single-cell gene expression data collected throughout *Caenorhabditis elegans* embryogenesis provides an excellent opportunity for investigating signaling networks underlying cell fate specification at systems, cellular and molecular levels.
Results: We propose a framework to infer signaling networks at cellular level by exploring the single-cell gene expression data. Through analyzing the expression data of *nhr-25*, a hypodermis-specific transcription factor, in every cells of both wild-type and mutant *C.elegans* embryos through RNAi against 55 genes, we have inferred a total of 23 genes that regulate (activate or inhibit) *nhr-25* expression in cell-specific fashion. We also infer the signaling pathways consisting of each of these genes and *nhr-25* based on a probabilistic graphical model for the selected five founder cells, 'ABarp', 'ABpla', 'ABpra', 'Caa' and 'Cpa', which express *nhr-25* and mostly develop into hypodermis. By integrating the inferred pathways, we reconstruct five signaling networks with one each for the five founder cells. Using RNAi gene knockdown as a validation method, the inferred networks are able to predict the effects of the knockdown genes. These signaling networks in the five founder cells are likely to ensure faithful hypodermis cell fate specification in *C.elegans* at cellular level.
Availability and Implementation: All source codes and data are available at the github repository https://github.com/xthuang226/Worm_Single_Cell_Data_and_Codes.git.
Contact: zhuyuan@cug.edu.cn
Supplementary information: Supplementary data are available at *Bioinformatics* online.

1 Introduction

Metazoan organism consists of different cell types, but their development is instructed by the same genome. At early stage of development, cell fate differentiation allows diversification of cell types. The fate of the early embryonic stem cells is differentiated into various cell types which is concurrent with cell division. Cell fate specification depends

on the expression of tissue-specific genes regulated by various signaling networks. However, the underlying signaling networks at cellular level in most organisms are not well understood.

It is possible, but time-consuming, to detect all the interactions and their effects among genes to reconstruct signaling networks through experiments, such as Chromatin ImmunoPrecipitation Sequencing (ChIP-Seq) and gene knockout. Therefore, it is necessary to develop computational approaches to infer the signaling networks using limited experiment data. Yeang *et al.* (2004) proposed a Physical Network Model based on Bayesian network theory, and used Boolean variables to describe the molecular networks. To take into account feedback loop detection, Bender *et al.* (2010) proposed an approach to reconstruct network using time course expression profiles and knockout data on the basis of a hidden Markov model. Ourfali *et al.* (2007) suggested an optimization model by using protein–protein interaction (PPI), protein–DNA interaction (PDI) and gene knockout data, named Signaling-regulatory Pathway INferencE (SPINE), which analyzed activating or inhibiting effect of each gene to explain maximum number of cause-effect pairs observed from gene knockout data. However, most of these methods were designed for the inference in yeast, a single-cell organism, and they mainly focused on edge orientation rather than inferring interaction effect. Without analysis of cellular level data, they are not suitable for inferring cellular signaling networks for cell fate specification in multicellular organism such as *C.elegans*.

Due to the availability of cellular level information on gene function, several methods and regulatory networks for single-cell type were proposed previously. Xu *et al.* (2014) focused on construction of a regulatory network responsible for self-renewal and pluripotency of mouse embryonic stem cells (mESCs). Yordanov *et al.* (2016) presented a methodology for automated reasoning a logical regulatory network model consistent with experimental observations, which was then applied to the analysis of mESCs. Both of their methods aim to refine a predefined regulatory networks based on experimental observations by considering logical functions for regulation. However, they are not suitable for inferring the potential intermediary genes and interactions to explain the indirect regulation effect. Stigler and Chamberlin (2012) inferred a regulatory network for muscle and skin development in *C.elegans* embryo, but only focused on C lineage and detected small number of cells. A recent research from Du *et al.* (2014) used single-cell gene expression data to investigate phenotypes changes of cells in 350-cell stage *C.elegans* embryo by RNA interference (RNAi) knockdown of twenty genes. They proposed an inferred mechanistic network involving both gene interaction and cell division. However, a detailed network with focus on a single-cell type remains to be seen.

In this article, a framework for cellular signaling network inference is proposed by using the time lapse single-cell gene expression data. It is applied for inferring signaling networks in different founder cells that mainly give rise to hypodermis (skin) in *C.elegans* embryo.

2 Materials and methods

We propose a framework to infer signaling networks at cellular level for cell fate specification based on multiple data types, including single-cell gene expression data, PPI, PDI and Genetic Interaction (GI) data. Here, we briefly sketch the framework in Figure 1. First, single-cell gene expression data both in wild-type and mutant are preprocessed and used to infer knockdown genes' effects at cellular level. Second, the inferred knockdown genes' effects along with PDI and GI data are formatted as cause-effect pairs to be subsequently used in pathway inference. After that, candidate paths are selected in a background network constructed by the PPI and PDI data. Then, the candidate paths and the cause-effect pairs are utilized to infer pathways with sign and direction for different founder cells. Finally, by integrating the inferred pathways, signaling networks at cellular level are reconstructed.

2.1 Data acquisition

The data exploited in this work consist of single-cell gene expression data and molecular interaction data of *C.elegans*. The single-cell gene expression data were obtained from an automated cell lineage tracing system, which captures images from a camera with fluorescence microscope to automatically trace *C.elegans* cells and measure fluorescence intensity of cell as gene expression during the *C.elegans* embryo development *in vivo* (see Ho *et al.*, 2015, Materials and Methods).

We selected *nhr-25* as a tissue-specific marker for hypodermis, whose gene expression was measured within 350-cell stage both in wild-type (16 replicates) and mutant (2–5 replicates for each single gene RNAi knockdown) *C.elegans* embryos. The single gene RNAi knockdown experiments were conducted for 53 genes to infer their effects on *nhr-25*, and additionally for *pop-1* and *lit-1* to validate the inferred network (see Supplementary Section I). These genes were chosen because

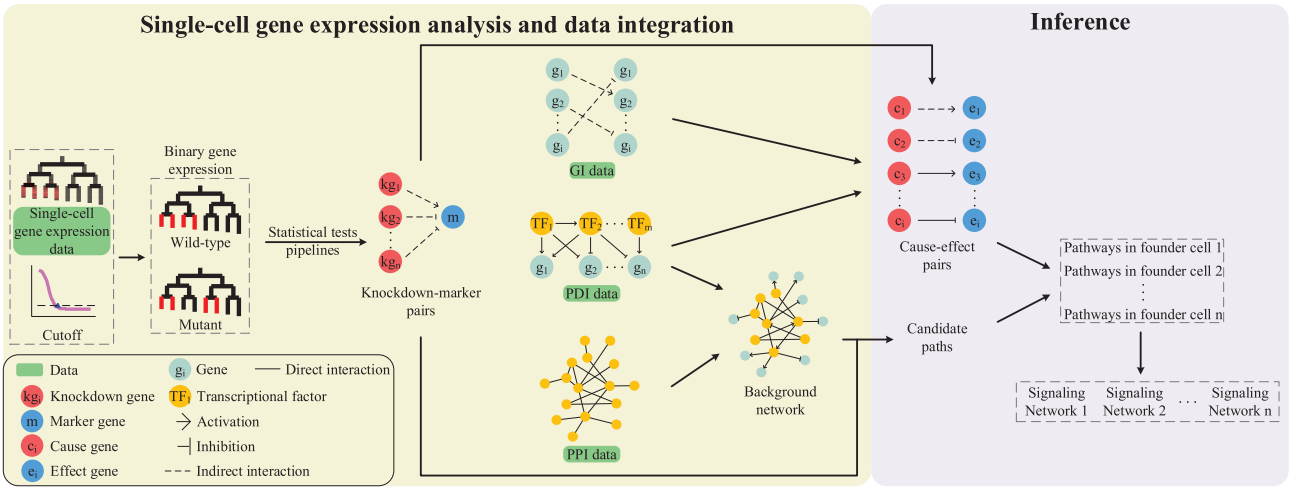

Fig. 1 The flowchart of signaling network inference based on single-cell gene expression data (Color version of this figure is available at *Bioinformatics* online.)

they were previously reported to be involved in hypodermis specification and development but their roles remain unknown. The details of single-cell gene expression measurement are described in Supplementary Section II. The single-cell *nhr-25* gene expression data of both wild type and mutant are available at the github repository.

To uncover signaling pathways from the knockdown genes to the marker, a physical molecular interaction network (background network) is imperative, which is reconstructed from PPI and PDI data (Yeang *et al.*, 2004; Zhang and Zhao, 2013). In this work, PPI data was obtained from an integrative weighted PPI network in *C.elegans*, which includes 5039 proteins involving with 12951 interactions (Huang *et al.*, 2016). 721 PDIs were obtained from Reece-Hoyes *et al.* (2011), which proposed an enhanced yeast one-hybrid assays for high-throughput gene-centered PDIs detection.

Similar to other methods (Dorn *et al.*, 2011; Ourfali *et al.*, 2007; Peleg *et al.*, 2010; Yeang *et al.*, 2004), cause-effect pairs, indicating that a gene holds an (activating or inhibiting) effect on another gene either directly or indirectly, were also included for the inference in this work, derived from three sources: (i) the inferred RNAi knockdown gene effect; (ii) gene regulatory effect data; (iii) genetic interaction data. The RNAi knockdown gene effects were inferred from wild-type and mutant single-cell gene expression data with methods in Section 2.2. 5611 gene regulatory effect data and 5187 genetic interaction data were retrieved from WormBase version WS248 with either activating (positive regulatory/enhancement) effect or inhibiting (negative regulatory/suppression) effect. All of them were formatted as cause–effect pairs. Totally, 7823 cause-effect pairs were obtained for the subsequent signaling pathway inference in Section 2.3.

2.2 Inference of knockdown gene effects in single cell and founder cell

After preprocessing single-cell gene expression data (see Supplementary Section III) and determining a reasonable cutoff (see Supplementary Section IV), cellular marker expression can be binarized and visualized on a cell lineage tree, shown as an example in a wild-type data set in Figure 2a. Among all the selected knockdown genes, some may not affect the marker expression, while others may hold either activating or inhibiting effect in different cells. For example, the difference of the marker expression between wild-type and *pop-1* gene knockdown data is shown in the bar chart in Figure 2b. The data points in the figure are averaged values. As *pop-1* gene is knocked down, the marker expression decreases in 'ABpla', 'ABpra' and 'Caa' lineages, and increases in 'ABarp' lineage, pointed by the green arrows in Figure 2b. Note that the knockdown gene may have different effects on the marker in different cells.

In Figure 2b, the changes of the marker expression between wild-type and mutant data set can be observed roughly. However, they are not able to quantitatively describe the knockdown gene effects at cellular level. To this end, wild-type and mutant expression data per cell are compared through a pipeline based on statistical hypothesis tests (See Figure 3a). Here, a nonparametric method, Wilcoxon rank-sum test (Mann–Whitney *U* test), is applied. For one cell, a statistical difference between wild-type and mutant expression data is first tested with a *P*-value. If not significant ($P > 0.05$), it indicates that the knockdown gene may not affect the marker expression in the cell. Further, if it is significantly different, the cutoff condition will be considered. Specifically, for the average of wild-type data and the average of mutant data, only one of them should be greater than the cutoff (determined by the method in Supplementary Section IV), while the other should be less than the cutoff. This cutoff condition can avoid two kinds of false-positive cases: (1) the marker expresses in both wild-type and mutant; (2) the marker expresses in neither wild-type nor mutant. We only consider those knockdown genes altering the marker expression status in the mutant cell. If the condition is met, then, two groups of data will be tested by Wilcoxon one-tailed test to detect which one is greater with a significant *P*-value. Finally, the test result indicates the

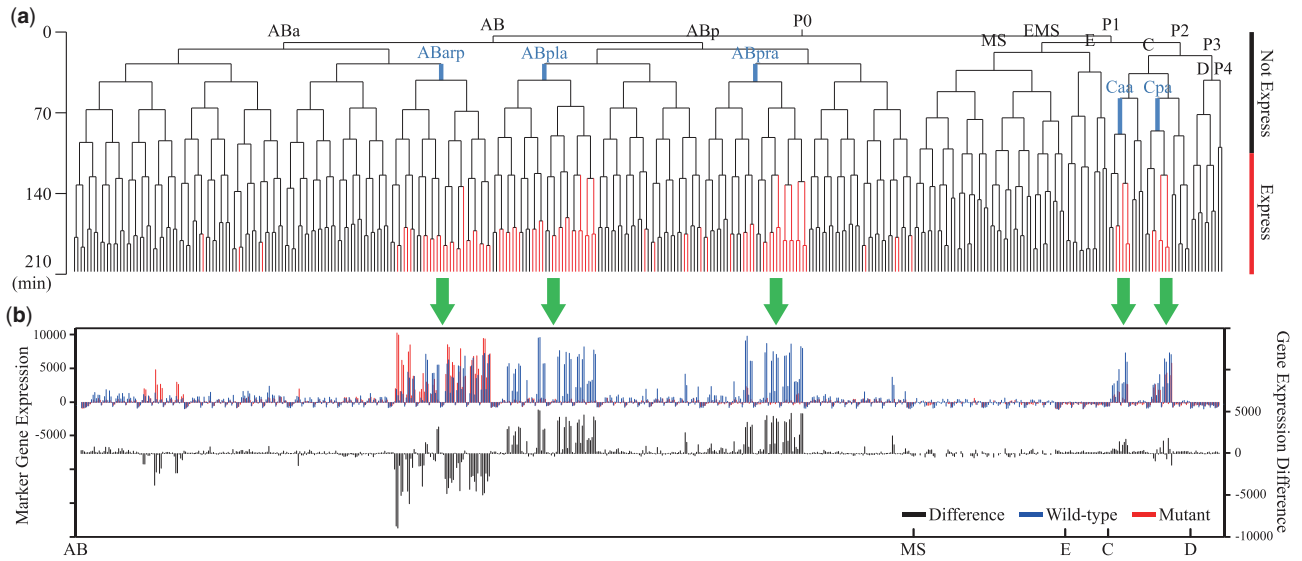

Fig. 2 Cell lineage tree and gene expression difference. (**a**) Binary *nhr-25* gene expression in cell lineage tree in one wild-type data set. Horizontal axis indicates all cells in the cell lineage, while vertical axis represents the cell division time. Red vertical lines represent the cells expressing *nhr-25*. Five founder cells, 'ABarp', 'ABpla', 'ABpra', 'Caa' and 'Cpa', are highlighted in blue, whose descendants express *nhr-25*. (**b**) *nhr-25* gene expression, which is averaged value, in wild-type and mutant (*pop-1* gene knockdown) in every cell. The horizontal axis indicates cells arranged from left to right corresponding to those in the cell lineage tree. Symbol 'AB' indicates a start in AB cell lineage. The cells on its right belong to AB cell lineage until the next symbol 'MS'. The MS, E, C and D cell lineages are depicted in the same manner. The upper bar chart, combining with the left vertical axis, shows *nhr-25* gene expression among different cells both in wild-type and the mutant data set. The lower bar chart, combining with the right vertical axis, shows the difference of *nhr-25* gene expression between wild-type and the mutant data set (wild-type data minus mutant data) (Color version of this figure is available at *Bioinformatics* online.)

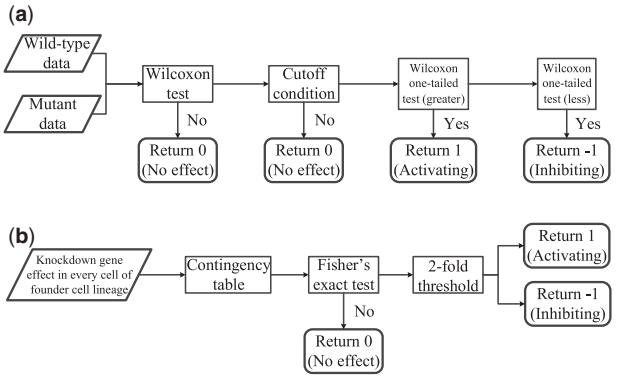

Fig. 3 Statistical hypothesis tests pipelines for knockdown gene effect inference. (**a**) In single cell. (**b**) In founder cell

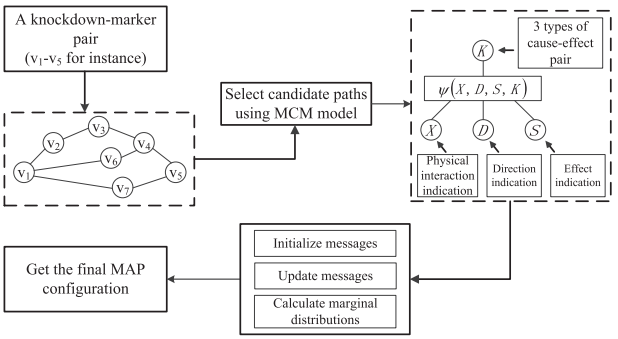

Fig. 4 The flowchart of pathway inference. An example of paths linking the knockdown gene v_1 and the marker v_5 to constitute a subnetwork is shown on the left. After selecting candidate paths, a factor graph, at the top right corner, illustrates the relationship of all random variables which define a pathway. By applying message passing algorithm, the final MAP configuration can be obtained

knockdown gene effect, either activation or inhibition, on the marker in the cell. For each knockdown gene, this pipeline was applied in every cell to infer its effects at cellular level.

During development, the cell fate, to generate specific cell type at later stage of the embryo, have been determined in the ancestral founder cells previously (Batra *et al.*, 2012). Therefore, it is necessary to detect the cell fate specification genes for individual founder cell. However, most of those genes express in descendants rather than in founder cells directly, such as *nhr-25* gene expressing in later stage in the cell lineage tree in Figure 2a. Therefore, it is rational to analyze data both in the founder cell and its descendants. In this work, five founder cells, 'ABarp', 'ABpla', 'ABpra', 'Caa' and 'Cpa', shown in blue in the top of the cell lineage tree in Figure 2a, were selected, because their descendants naturally develop into hypodermis (Du *et al.*, 2014). The inferred knockdown gene effects in single cell are used for inferring its effects in founder cells by a statistical hypothesis tests pipeline in Figure 3b. First, a 2×2 contingency table is constructed, with two columns representing the number of cells harboring significant effect and no effect respectively for a specific knockdown gene, while with two rows in wild-type and mutant respectively. Then, Fisher's exact test is performed for this contingency table to test whether this knockdown gene has a significant effect on the marker in this founder cell. If significant ($P < 0.1$), a 2-fold threshold is applied for deciding whether it is an activating or inhibiting effect. In other words, under the condition of significant *P*-value, if the number of activating effect cells is larger than 2-fold of inhibiting number, this knockdown gene may have an integrally activating effect on the marker in the founder cell, and vice versa. The inferred knockdown gene effect is formatted as cause–effect pair with a sign ($\pm$) indicating activation or inhibition. By applying this pipeline, effects of all knockdown genes were inferred in different founder cells for the marker.

2.3 Pathway inference

The knockdown gene effects can be inferred at cellular level by the methods in Section 2.2. However, the knockdown gene may not affect the marker expression directly. It is reasonably assumed that several intermediary genes/proteins pass the regulatory signals from the knockdown gene to the marker. In this section, we propose a method to infer signaling pathways with direction and sign from the knockdown gene to the marker.

2.3.1 Candidate path selection

To infer signaling pathway, a weighted physical molecular interaction network consisting of PPIs and PDIs, as a background network, is necessary for finding possible connecting paths from the knockdown gene to the marker (Gitter *et al.*, 2011). However, it is Non-deterministic Polynomial-time hard (NP-hard) to enumerate all possible connecting paths in a complex network. Therefore, an algorithm for candidate path selection is proposed based on the Markov Chain Model (MCM). All paths less than a specified length L are enumerated from the knockdown (cause) gene v_c to the marker (effect) gene v_e. Thus, all of them generate a subnetwork, in which nodes can be considered as a sequence of random variables in a Markov chain, namely, the present node is independent with the future and past nodes. Taken Figure 4 as an example, $P(v_5|v_1, v_2, \ldots, v_4) = P(v_5|v_4)$. Here, the probability is calculated as $P(v_i|v_{i-1}) = r_{i-1,i}$, where $r_{i-1,i}$ is the RSPGM score in Huang *et al.* (2016) if the interaction between v_{i-1} and v_i is a PPI, while $r_{i-1,i}$ is the reliability assessed by Reece-Hoyes *et al.* (2011) if a PDI. Besides, all the probability should be satisfied by $\sum_{v_c \in N_{v_i}} P(v_c|v_i) = 1$, where N_{v_i} includes all neighbors of v_i. Therefore, for any path from v_c to v_e and a specified length L, the path score $S_{v_c \to v_e}$ can be calculated as follow,

$$S_{v_c \to v_e} = P(v_{c+1}|v_c)P(v_{c+2}|v_{c+1}) \cdots P(v_e|v_{c+L-1}). \quad (1)$$

After that, the candidate paths are selected by filtering the paths through the following threshold of path score with parameter $0 \leq \alpha \leq 1$,

$$S_{v_c \to v_e} \geq \alpha \max_{\text{Path}} S_{v_c \to v_e}, \quad (2)$$

where $\text{Path} = \{(v_c \to v_{c+1} \to \cdots \to v_e) | v_{c+1}, v_{c+2}, \ldots, \in G\}$. If setting $\alpha = 1$, the path with the highest path score is selected as the candidate path.

2.3.2 Define variables to describe a signaling pathway

In Section 2.3.1, the method only selects candidate paths. However, directions and signs of the edges are still not determined. Here, several variables, V, E, X, D, S and K, are defined to describe the properties of a signaling pathway. $V = \{v_i\}$ is a collection of all genes/proteins on the pathway. $E = \{e_{i(i+1)}\}$ is a collection of all physical interactions on the pathway. $X = \{x_{i(i+1)}\}$, where $x_{i(i+1)} = 1$ if a physical interaction exists between node v_i and node v_{i+1} else $x_{i(i+1)} = 0$. $D = \{d_{i(i+1)}\}$, where $d_{i(i+1)} = 1$ if the direction of the physical interaction is consistent with the pathway direction (the knockdown gene $\to$ the marker) else $d_{i(i+1)} = 0$. $S = \{s_{i(i+1)}\}$, where $s_{i(i+1)} = 1$ if an activating effect, $s_{i(i+1)} = -1$ if an inhibiting effect,

and $s_{i(i+1)} = 0$ if no effect. $K = \{k_{mn}, m < n\}$ are three types of cause-effect pair in Section 2.1, demonstrating the (activating or inhibiting) effect between any two nodes v_m and v_n. The amount of variables in X, D, S depends on the length of the candidate path. For example, if there are two intermediary genes/proteins between the knockdown gene and the marker on the candidate path, the V contains four nodes (v_1, v_2, v_3, v_4) and the E contains three interactions (e_{12}, e_{23}, e_{34}). Therefore, totally nine variables, (x_{12}, x_{23}, x_{34}, d_{12}, d_{23}, d_{34}, s_{12}, s_{23}, s_{34}) in X, D, S, need to be inferred to determine the pathway.

2.3.3 Reasonable signaling pathway to explain the knockdown effect

There are several configurations for $x_{i(i+1)} \in X$, $d_{i(i+1)} \in D$ and $s_{i(i+1)} \in S$. However, only partial configurations can explain the regulatory effect from the knockdown gene to the marker. A reasonable signaling pathway should comply with the following conditions: (1) all interactions are present; (2) the directions of all interactions are consistent with the pathway direction; (3) the effects of interactions explain the cause-effect pairs on the pathway. To encode these conditions, a potential function $\psi(X, D, S, K)$ with X, D, S and K is expressed as follow,

$$\psi(X, D, S, K) = \prod_{x_{i(i+1)} \in X} x_{i(i+1)} \cdot \prod_{d_{i(i+1)} \in D} d_{i(i+1)} \cdot \prod_{k_{mn} \in K} I\left(\prod_{m \le i \le (n-1)} s_{i(i+1)} = k_{mn} \right), \quad (3)$$

where $I(\cdot)$ is the indicator function. This potential function returns 1 for the reasonable pathway and 0 for the unreasonable pathway, which maps each configuration into a binary value pertaining to the constraints from the observed data.

2.3.4 Probabilistic graphical model for inference

To find the most likely configuration of variables describing a reasonable pathway to explain the cause-effect pairs, a factorization of the joint distribution of all random variables is defined as follow,

$$P(X, D, S|K) \propto P(X) \cdot P(D) \cdot P(S) \cdot \psi(X, D, S, K). \quad (4)$$

The joint distribution of X, D and S is proportional to the multiplication of the potential function in Equation (3) and the marginal distribution of each variable in X, D and S based on a factor graph at the top right corner of Figure 4.

The goal of pathway inference is to determine the intermediary nodes and their interactions with directions and signs. In other words, it needs to find the most likely configuration, Maximum A Posteriori (MAP) of the joint distribution, of all related random variables. In Equation (4) and Figure 4, to obtain the marginal distribution of each variable, a message passing algorithm, sum-product algorithm, can be applied in the factor graph for the inference (Yeang *et al.*, 2004). Due to a tree structure in the factor graph in Figure 4, the inference steps are illustrated as: (1) Initialize all messages from the variables to the potential function based on the data involved in the pathway; (2) Update all messages from the potential function to the variables; (3) Compute marginal distribution of each variable; (4) Compute MAP of the joint distribution in Equation (4) to find the MAP configuration as the inferred reasonable pathway. By applying this method, the existence, directions and signs of interactions on the candidate path can be inferred.

3 Results

3.1 The inferred knockdown gene effects in different founder cells

Based on the wild-type single-cell *nhr-25* expression data, a reasonable cutoff was determined as 1466.8 which is close to the result (1274.41) in Du *et al.* (2014). The difference may be due to the different fluorescence microscope laser setting in the automated cell lineage tracing system.

To detect which genes involve *nhr-25* gene expression regulation, we conducted RNAi single gene knockdown experiments. As mentioned in Section 2.1, 53 genes as well as *pop-1* and *lit-1* were knocked down. Through methods in Section 2.2, combining with the determined cutoff, the effects of these 55 genes were inferred in each founder cell. We found that 23 knockdown genes possess either activating or inhibiting effect on *nhr-25* in different founder cells, shown in Table 1. All inferred effects have a significant P-value ($P < 0.1$), except for the activating effect of *lit-1* in founder cell 'Caa' ($P = 0.220$) and the inhibiting effect of *pop-1* in founder cell 'ABarp' ($p = 0.119$). We consider that these two inferred effects are reliable, because when we use data beyond 350-cell stage, the P values are significant. This suggests that the knockdown effects of *lit-1* and *pop-1* may be observed in late stage of the embryo. The inferred effects of 21 genes (in Table 1 except for *lit-1* and *pop-1*) were used for networks inference, while the inferred effects of *lit-1* and *pop-1* (marked with '⋆' in Table 1) were used to validate the inferred networks, described in Section 3.3.

The genes in Table 1 are closely related to hypodermis cell fate specification. Among them, several genes have been known for the importance of hypodermis formation, such as *elt-1*, *elt-3*, *nhr-25*, *pal-1*, *tbx-8* and *tbx-9* (Du *et al.*, 2014). In our result, *elt-1*, *elt-3*,

Table 1 Effects of 23 knockdown genes on *nhr-25* in the five founder cells

Gene	ABarp	ABpla	ABpra	Caa	Cpa
B0205.6		−	−		
abce-1	−	−			
cbp-1	+		+	+	+
cogc-2	−	−	−		
elt-1		+			
elt-3	−				
gad-1	+	+	+	+	
hlh-2			−		
isw-1	−	−	−		
lag-1		+	+		
let-526	+	+	+	+	+
lim-7			−		
lit-1⋆	+	+	+	+	
nhr-25	+	+	+	+	+
nhr-68	−				
nol-5	−	−	−		
pal-1	−			+	+
pes-1		+			
pop-1⋆	−	+	+	+	
skn-1			−		
tads-1	+	+	+	+	+
tbx-33		+	+		
tbx-37+38	−				

Rows represent knockdown genes, while columns represent founder cells. '+' and '−' represent activating and inhibiting effect from the knockdown gene to the marker, respectively. '⋆' represents that the effects of the gene will be used for signaling network inference validation.

nhr-25 and *pal-1* also have effect on *nhr-25*. Moreover, *tbx-8* and *tbx-9*, as T-box transcription factors, are redundant pairs of *tbx-37* and *tbx-38* (Mango, 2007), which are also contained in Table 1. Therefore, our results include these six known hypodermis formation genes as well as other seventeen newly discovered genes.

Note that it is intriguing that one knockdown gene has an activating effect in some cells but an inhibiting effect in others. Similar to Shao *et al.* (2013), we found that *lag-1*, *pop-1* and *pal-1* function in this way (see Supplementary Table S1). Particularly, they hold activating effect in anterior cells, while inhibiting effect in posterior cells. This suggests that the differential regulation effects for a knockdown gene in different cells may be due to asymmetry cell division between anterior and posterior cells.

3.2 The inferred signaling networks in different founder cells

The inferred effect only indicates the global effect from the knockdown gene to the marker in each founder cell. To select candidate connecting paths, we used the setting of $L = 5$ which is consistent with Gitter *et al.* (2011), because (1) short signaling cascades are reasonable; (2) average pathway length is five in KEGG database and the Science Signaling Database of Cell Signaling. Moreover, we used the setting of $\alpha = 1$ to select the path with highest score. The intermediary genes/proteins and their interactions with direction and sign were then inferred through the pathway inference method (see Section 2.3). Finally, in each founder cell, the inferred pathways, from all effective knockdown genes to the marker, were integrated to construct a signaling network regulating *nhr-25* expression in the founder cell. The inferred signaling networks for the five founder cells are shown in Fig. 5. In the networks, effects of many interactions are not determinate due to the limiting cause-effect pairs in Section 2.1. Although the signs of interactions are partially determined, our method is still able to infer the intermediary genes/proteins and the directions of interactions.

To measure the difference of the inferred networks, we exploited three network distance metrics, Hamming Distance (HD) (Hamming, 1950), Mean Absolute Difference (MAD) (Butts and Carley, 2001) and Generalized Hamming Distance (GHD) (Ruan *et al.*, 2015). These three types of distances among networks for the five founder cells were computed with *P*-values (see Supplementary Table S2). In the result, most of these inferred networks are not significantly different. The slight differences could arise from partially localized re-wiring under different conditions, such as extracellular signaling. However, in this paper, the experiments were not intended to capture any extracellular signaling, but focused on inferring networks for individual founder cells with fate of hypodermis.

3.3 Validation of the inferred signaling networks

To validate whether the inferred signaling networks are reasonable, we intended to investigate their availability of explaining perturbation experiment result. Particularly, we focused on *pop-1* and *lit-1*, because they are hub nodes in the inferred networks. We intended to validate whether the prediction of *pop-1* and *lit-1* effects in the inferred networks are consistent with experimental knockdown effects.

In networks of 'ABarp', 'ABpla', 'ABpra' and 'Caa', all interactions from *lit-1* to *nhr-25* are activation (see Fig. 5a–d). Therefore, we predicted that the global effect of *lit-1* on *nhr-25* would be activating in these four founder cells. Similarly, we predicted that the global effect of *pop-1* on *nhr-25* would be activating in 'ABpla',

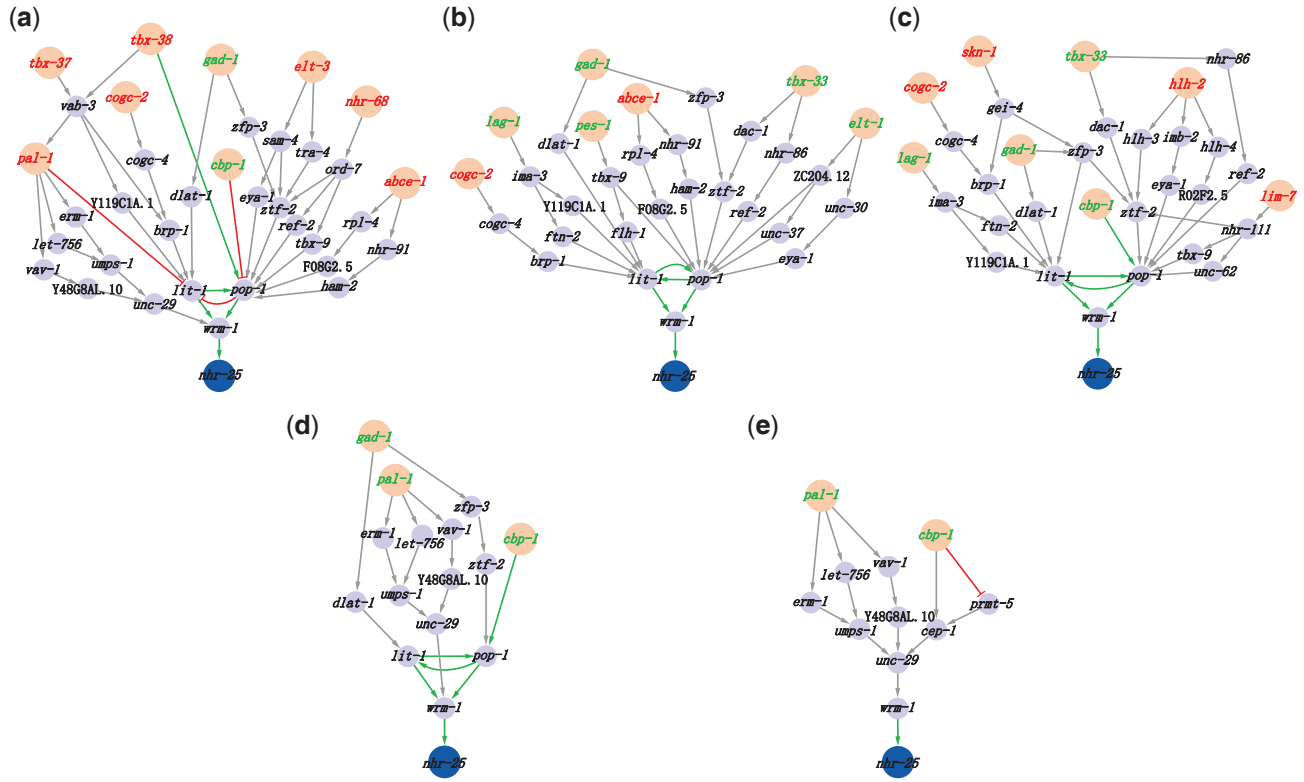

Fig. 5 The inferred signaling networks for *nhr-25* in the five founder cells. (**a**) In 'ABarp'. (**b**) In 'ABpla'. (**c**) In 'ABpra'. (**d**) In 'Caa'. (**e**) In 'Cpa'. Green and red arrows represent activating and inhibiting effect respectively. Gray arrow represents that the effect cannot be determined. Dark blue node represents the marker. Orange node represents the knockdown gene. Green and red labels in the knockdown gene represent its activating and inhibiting effect on *nhr-25,* respectively (Color version of this figure is available at *Bioinformatics* online.)

'ABpra' and 'Caa', but inhibiting in 'ABarp'. This predicted inhibiting effect is due to an inhibiting interaction from *pop-1* to *lit-1* in 'ABarp' network, indicating that $pop\text{-}1 \xrightarrow{-} lit\text{-}1 \xrightarrow{+} wrm\text{-}1 \xrightarrow{+} nhr\text{-}25$ could be a reasonable pathway to regulate *nhr-25* (see Figure 5a).

To validate above predictions, we performed both *pop-1* and *lit-1* single gene RNAi knockdown experiments. Based on the knockdown data, through the methods in Section 2.2, *pop-1* was inferred as activating effect in 'ABpla', 'ABpra' and 'Caa', but inhibiting effect in 'ABarp', while *lit-1* was inferred as activating effect in all these four founder cells, shown in Table 1 marked with '⋆'. Therefore, the predictions are consistent with the knockdown experiment results, indicating that the inferred cellular networks are reasonable for regulating *nhr-25* expression.

In the inferred networks, we found that a common module consisting of *lit-1*, *pop-1* and *wrm-1*, which are the components of Wnt signaling pathway (KEGG PATHWAY: cel04310). This suggests that Wnt signaling pathway involves regulating *nhr-25* gene expression. Coincidently, Hajduskova *et al.* (2009) and Asahina *et al.* (2006) have found that *nhr-25* cooperates with Wnt signaling pathway for cell fate specification in *C.elegans*. Consequently, our inferred cellular signaling networks are reasonable for hypodermis cell fate specification in *C.elegans*.

4 Discussion

The underlying signaling networks of cell fate specification are responsible for maintaining a normal cell differentiation progression for tissue formation. From the inference by the proposed framework exploiting single-cell *nhr-25* gene expression data, a list of 23 genes has been identified for hypodermis cell fate specification. These genes differentially affect *nhr-25* gene expression in different founder cells and their descendants to determine their fate. By inferring the latent regulatory signaling networks, it reveals a regular and precise development of hypodermis in normal organism of *C.elegans*.

Through the proposed framework, we inferred signaling networks in the selected five founder cells, because most of their descendants have fate of hypodermis development. Note that, due to the strategy, this framework can be used to infer networks for any specified cells, such as the descendants of the five founder cells or the cells without fate of hypodermis. However, if different cells have the same set of effective knockdown genes, which lead to the same inferred pathways, the resultant networks will be the same.

Based on the framework strategy, we briefly discuss several factors influencing the inferred pathways which in turn determine the resultant networks. First of all, the set of effective knockdown genes can determine upstream of the inferred pathways. Second, because the pathways are extracted from a background network, different PPI and PDI data can influence the inferred pathways. Moreover, the length setting and the α setting in path extraction step in Section 2.3.1 can affect the length and the number of the inferred pathways. Finally, cause-effect pairs can influence the regulatory effect of interactions in the inferred pathways. Therefore, due to these factors, reliable data supporting and appropriate parameter setting will lead to reasonable and detailed networks.

It can be observed that several parallel pathways existing in the inferred networks. This is reasonable that overlapping parallel pathways link the cause and effect genes, similar to Piloto *et al.* (2007) and Gitter *et al.* (2011). Moreover, the signs of interactions in the inferred networks are partially determined, because limited cause-effect pairs at this stage. However, we have already demonstrated that these networks are reasonable to reveal regulatory mechanisms for hypodermis cell fate specification in *C.elegans*. In the future, complete inferred signaling networks could be accomplished when more cause-effect pairs from the PDI data, GI data, or others become available.

Funding

This work was supported by the National Natural Science Foundation of China (Project 11401110, 61532014 and 91530113), the Hubei Provincial Natural Science Foundation of China under Grant 2015CFA010, the Research Center Foundation of School of Automation of China University of Geosciences (Wuhan) (Project AU2015CJ008), the Fundamental Research Funds for the Central Universities, China University of Geosciences (Wuhan), Natural Science Foundation of Guangdong province (Project 2013KJCX0086), Hong Kong Research Grants Council (Project HKBU5/CRF/11G), Project CityU 11214814 and the China Postdoctoral Science Foundation funded project (2016M600769).

Conflict of Interest: none declared.

References

Asahina,M. *et al.* (2006) Crosstalk between a nuclear receptor and beta-catenin signaling decides cell fates in the *C. elegans* somatic gonad. *Dev. Cell*, **11**, 203–211.

Batra,R. *et al.* (2012) Time-lapse imaging of neuroblastoma cells to determine cell fate upon gene knockdown. *PLoS ONE*, **7**, e50988.

Bender,C. *et al.* (2010) Dynamic deterministic effects propagation networks: learning signalling pathways from longitudinal protein array data. *Bioinformatics*, **26**, i596–i602.

Butts,C.T., and Carley,K.M. (2001). Multivariate methods for inter-structural analysis. *CASOS Working Paper*: Carnegie Mellon University, Pittsburgh, PA: Center for the Computational Analysis of Social and Organization Systems.

Dorn,B. *et al.* (2011) Exploiting bounded signal flow for graph orientation based on cause-effect pairs. *Algorithms Mol. Biol.*, **6**, 21.

Du,Z. *et al.* (2014) De novo inference of systems-level mechanistic models of development from live-imaging-based phenotype analysis. *Cell*, **156**, 359–372.

Gitter,A. *et al.* (2011) Discovering pathways by orienting edges in protein interaction networks. *Nucleic Acids Res.*, **39**, e22.

Hajduskova,M. *et al.* (2009) The nuclear receptor NHR-25 cooperates with the Wnt/beta-catenin asymmetry pathway to control differentiation of the t seam cell in *C. elegans*. *J. Cell Sci.*, **122**, 3051–3060.

Hamming,R.W. (1950) Error detecting and error correcting codes. *Bell Syst. Tech. J*, **29**, 147–160.

Ho,V.W. *et al.* (2015) Systems-level quantification of division timing reveals a common genetic architecture controlling asynchrony and fate asymmetry. *Mol. Syst. Biol.*, **11**, 814.

Huang,X.T. *et al.* (2016) An integrative *C. elegans* protein-protein interaction network with reliability assessment based on a probabilistic graphical model. *Mol. Biosyst.*, **12**, 85–92.

Mango,S.E. (2007). The *C. elegans* pharynx: a model for organogenesis. *WormBook*, pages 1–26.

Ourfali,O. *et al.* (2007) SPINE: a framework for signaling-regulatory pathway inference from cause-effect experiments. *Bioinformatics*, **23**, i359–i366.

Peleg,T. *et al.* (2010) Network-free inference of knockout effects in yeast. *PLoS Comput. Biol.*, **6**, e1000635.

Piloto,O. *et al.* (2007) Prolonged exposure to FLT3 inhibitors leads to resistance via activation of parallel signaling pathways. *Blood*, **109**, 1643–1652.

Reece-Hoyes,J.S. *et al.* (2011) Enhanced yeast one-hybrid assays for high-throughput gene-centered regulatory network mapping. *Nat. Methods*, **8**, 1059–1064.

Ruan,D. *et al.* (2015) Differential analysis of biological networks. *BMC Bioinformatics*, **16**, 327.

Shao,J. *et al.* (2013) Collaborative regulation of development but independent control of metabolism by two epidermis-specific transcription factors in *Caenorhabditis elegans*. *J. Biol. Chem.*, **288**, 33411–33426.

Stigler,B., and Chamberlin,H.M. (2012) A regulatory network modeled from wild-type gene expression data guides functional predictions in *Caenorhabditis elegans* development. *BMC Syst. Biol.*, **6**, 77.

Xu,H. *et al.* (2014) Construction and validation of a regulatory network for pluripotency and self-renewal of mouse embryonic stem cells. *PLoS Comput. Biol.*, **10**, e1003777.

Yeang,C.H. *et al.* (2004) Physical network models. *J. Comput. Biol.*, **11**, 243–262.

Yordanov,B. *et al.* (2016) A method to identify and analyze biological programs through automated reasoning. *Npj Syst. Biol. Appl.*, **2**, 16010.

Zhang,X.D., and Zhao,X.M. (2013) Computational approaches for identifying signaling pathways from molecular interaction networks. *Curr. Bioinform.*, **8**, 56–62.