Toward deciphering developmental patterning with deep neural network

Jingxiang Shen[1], Mariela D. Petkova[4,5,6], Feng Liu[1,2], Chao Tang[1,2,3,*]

[1]Center for Quantitative Biology, [2]School of Physics, [3]Peking-Tsinghua Center for Life Sciences, Peking University, Beijing 100871, China
[4]Joseph Henry Laboratories of Physics and [5]Lewis-Sigler Institute for Integrative Genomics, Princeton NJ 08544, United States
[6]Program in Biophysics, Harvard University, Cambridge MA 02138, United States
(Date: July, 2018)

* Email: tangc@pku.edu.cn

Abstract

Dynamics of complex biological systems is driven by intricate networks, the current knowledge of which are often incomplete. The traditional systems biology modeling usually implements an *ad hoc* fixed set of differential equations with predefined function forms. Such an approach often suffers from overfitting or underfitting and thus inadequate predictive power, especially when dealing with systems of high complexity. This problem could be overcome by deep neuron network (DNN). Choosing pattern formation of the gap genes in *Drosophila* early embryogenesis as an example, we established a differential equation model whose synthesis term is expressed as a DNN. The model yields perfect fitting and impressively accurate predictions on mutant patterns. We further mapped the trained DNN into a simplified conventional regulation network, which is consistent with the existing body of knowledge. The DNN model could lay a foundation of "*in-silico-embryo*", which can regenerate a great variety of interesting phenomena, and on which one can perform all kinds of perturbations to discover underlying mechanisms. This approach can be readily applied to a variety of complex biological systems.

Introduction

The early embryogenesis of *Drosophila* is a well-studied model system in developmental biology, characterized by a rapid cascade of gene expression patterns[1]. Under the guidance of maternal effect morphogens, a handful of gap genes form sophisticated spatial patterns across the embryo, serving as the blueprint for future body plan. Large amounts of experimental and modeling efforts have been devoted to uncovering the genetic interaction network and regulatory mechanisms underlying the pattern formation[2-12], but mysteries still remain[2,3,13].

Various mathematical models of gap genes' expression have been constructed[6-9]. One kind of model[8], as in most modeling approaches of biological systems, starts with a presumed network inferred from a body of experimental work, and/or simplified by the author's opinion of what is important. Differential equations describing the rate change of each gene expression are written down, with gene regulation modeled by specific mathematical functions, e.g. the Hill function. Recognizing the fact that knowledge on regulations of gap genes may not be complete, another kind of model adopts a reverse engineering approach[6,7]. Genetic interactions are effectively expressed as a single layer neural-network-like architecture, with no prior constrains on regulatory structures. Regulations then emerge from data fitting. Both kinds of modeling had considerable success: certain important phenomena can be explained, gene expression data fitted, and the emerged regulation network in the second approach made some sense in comparing with the known knowledge. However, these models have inadequate predictive power.

This weakness in predictive power is natural here. The complexity in real biological systems such as this one may well exceed the capacity of these kinds of models. For example, the expression of each gap gene is contributed by 2~5 regulatory modules[9,14] (enhancers and shadow enhancers[15,16]), each of which is regulated differently, and

dynamical switch could happen between different enhancers[15]. Within each module, the regulatory sequence usually bears 10~20 binding sites of different transcription factors[17] with unknown cooperativity among them[9,18]. Furthermore, apart from the 4 gap genes (*hunchback* (*hb*), *Krüppel* (*Kr*), *knirps* (*kni*) and *giant* (*gt*)) focused by most models and quantitative experiments, there are very likely to be a number of other genes relevant to this process as suggested by bioinformatics search, or even unknown factors as suggested by the experiment[14]. These and other unknown complexity may introduce strong nonlinearity within the equivalent regulation functions, making it almost impossible to be expressed by predefined formulas.

This sort of dilemma is not uncommon when dealing with complex systems. On the one hand, we would like to simplify the system, but often have little idea how to simplify it or whether it can in principle be simplified -- the models may easily be oversimplified. On the other hand, even if one manages to obtain equations with enough complexity, they typically contain too many parameters to avoid overfitting with finite amount of data. In some cases, this problem could be alleviated by a recently developed adaptive modeling approach for dynamical systems[19]. But its applicability in more demanding situations, such as the spatiotemporal patterning here, has yet to be tested.

In this study, we try a different approach to this complex problem -- deep neural networks (DNN)[20-22]. We hope DNN, instead of regulation equations with prefixed forms, can alleviate the dilemma of model capacity. For reasons not yet completely clear, neural networks have almost infinite fitting power, but hardly overfit even without any regularization techniques[23]. To a certain extent, it is a kind of "self-adapting model", adjusting its own capacity to fit and avoid overfitting, thus overcoming the above-mentioned difficulty of traditional equation-based models. In a sense, our approach could be viewed as an upgraded version of the gene circuit models[24], but motivation and thus results are different: instead of seeking directly for

a unique regulation network with prefixed regulation forms, we aim to mimic this complex system as accurately as possible at the expense of using a black box. The DNN model is then validated with predictions on mutants' patterns, and can, in principle, be used as an "*in-silico-embryo*" on which we can perform all kinds of perturbations, so as to discover possible underlying mechanisms in such an indirect manner.

Results

Model Setup

As the *Drosophila* embryo is at the syncytial stage when gap gene patterns form (12-14th nucleus cycle (nc)), the spatiotemporal dynamics of these expressions could in theory be described by equations with synthesis, degradation and diffusion terms. Since the diffusion constants of gap proteins are estimated to be 1 $\mu m^2/s$ (around 10% embryo length within an hour)[4,8,24], we neglect diffusion for simplicity (including diffusion does not improve the performance). The dynamic equations are:

$$\frac{\partial g_i(x,t)}{\partial t} = F_i(g_1, g_2, g_3, g_4, m_1, m_2, m_3) - \frac{1}{\tau} g_i(x,t) \qquad (1)$$

Here, $g_i(x,t)$ stands for the expression level of gap gene i at spatial grid number x and time step t. Four gap genes, *hb*, *Kr*, *kni* and *gt*, are considered here. $m_i(x)$ denotes maternal morphogens, which are viewed as stable inputs throughout the relevant time period. The degradation rates for all 4 gap genes are set to be the same, as a trainable parameter; thus all the regulations should be contained in the synthesis term F_i, which gives the synthesis rate for gap gene i out of the current local expression level of the gap genes and maternal morphogens. With no prior knowledge on the regulation network nor the functional form of F_i being assumed, we use a 4-layer fully connected neural network to simulate F_i (i=1,2,3,4). Solving Eq. 1 numerically is then equivalent to recurrent architecture with F as the recurrent block (Fig. 1).

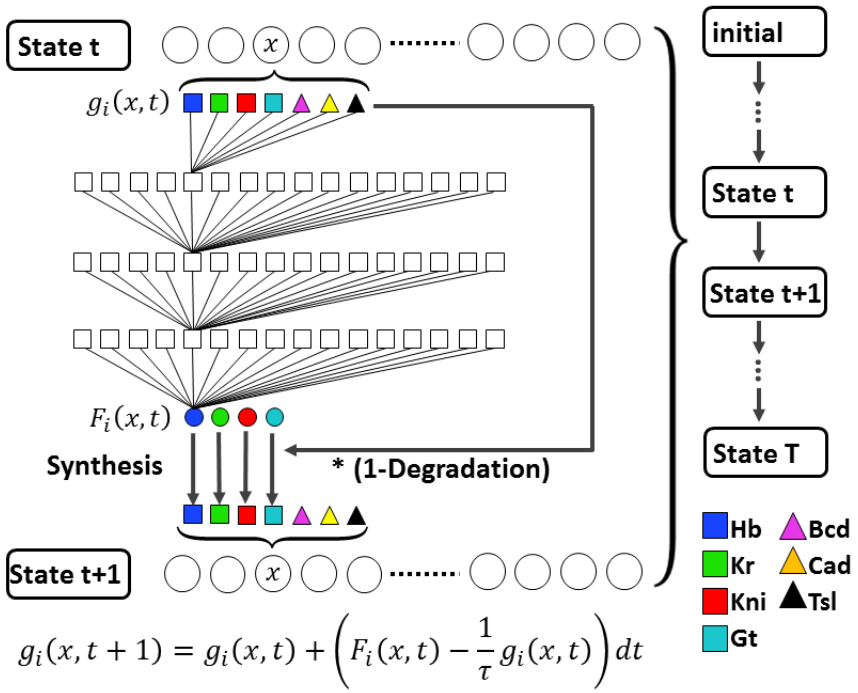

Figure 1. Architecture of the DNN model. Within the recurrent block, synthesis term is represented as a fully connected neural network with seven inputs, including four gap genes (colored rectangles) and three maternal inputs (colored triangles), and four outputs (synthesis rates of the four gap genes). Output patterns are then calculated by the recurrent network.

Maternal factors Bicoid (Bcd), Caudal (Cad) and Torso-like (Tsl) are selected as explicit maternal input patterns[1]. Among them, Cad pattern is assumed to be uniquely determined by Bcd as suggested by both biological knowledge and most previous models (Eq. S1)[25]. Another important maternal effect gene, Nanos (Nos), is assumed to take effect purely by shaping the initial condition of Hb[26,27]. The other gap genes all start with zero initial conditions (see Supplementary Information S1 for details).

Loss function for training is set to be the Euclidean distance between a selected set of experiment data $G_i(x,t)$ and the corresponding model pattern $g_i(x,t)$, for wild type (wt) and/or mutant systems (mut).

$$\mathrm{Loss}_{\mathrm{wt/mut}} = \sum_t \sum_i \sum_x \left(g_i(x,t) - G_i(x,t)\right)^2$$

$$\mathrm{Loss} = \sqrt{\mathrm{Loss}_{\mathrm{wt}} + \sum \mathrm{Loss}_{\mathrm{mut}}} \tag{2}$$

Thus, network F (synthesis term) is trained to form the desired patterns from given initial conditions and maternal inputs (see Supplementary Information S3 for details).

Training

Overfitting could be a problem, as our DNN model has about 750 parameters but the quantitative dataset we can collect are only dozens of frames. Moreover, avoiding overfitting can be more demanding in this study: unlike typical deep learning tasks where test and training data are sampled from the same distribution and in most cases features in test data are completely reflected in the training set, here wild type and mutants *are* different.

Based on multiple trials, the model achieves the best performance if the training set only uses 7 frames of the wt gap gene expression time course data (~5.5 min a frame, 8 to 41 minutes in nc14)[5] together with a snapshot of maternal factor triple mutant (Bcd^-;Nos^-,Tsl^-) at around 40 minutes in nc14 (Fig. S4B)[28]. Unsurprisingly, due to the powerful fitting capacity of DNN model, lots of details within the data can be well fitted (Fig. 2); e.g., relative heights of different peaks, the anterior peak of Kni, the posterior peak of Hb, and the dynamic anterior shift of abdomen patterns. Also, the training converges rather fast computationally: within only a couple of minutes on a common desktop.

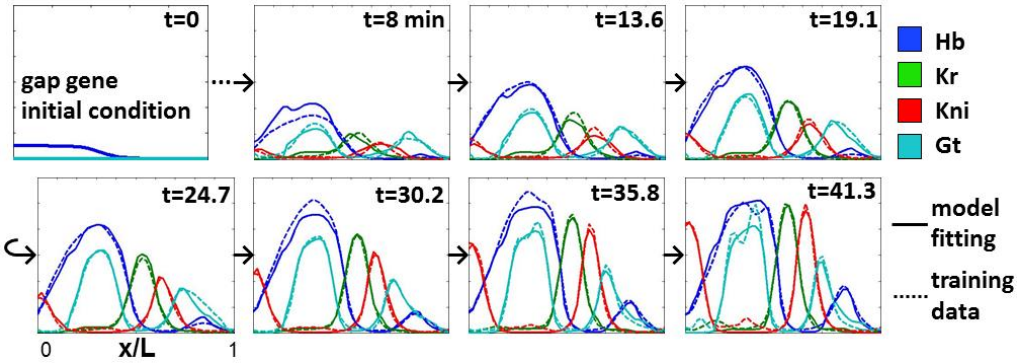

Figure 2. Fitting results of the training data. The fitting results (solid lines) of the expression profiles of Hb (blue), Kr (green), Kni (red) and Gt (cyan) along the A-P axis are very close to the training data (dashed line). Training data is seven frames of wt data and one frame of maternal morphogen triple mutant data, as described in the text.

Prediction

Surprisingly, the trained DNN model yields excellent predictions on the gap gene profiles in almost all the mutants, including various double mutants[28]. The number, position and even the relative intensity of almost all the peaks in the gap gene profiles are well predicted (Fig. 3). Interestingly, some delicate details are also captured by the prediction: (1) in Tsl^- and Tsl^-;Nos^- mutants, the height of anterior Kni peak drops half compared with wt; (2) in Bcd^-;Nos^- mutant, two symmetric small peaks of Gt exists; (3) in Kr^- mutant, Kni peak changes position and lies under Gt peak; (4) when Bcd dosage is halved or doubled (Bcd1X or Bcd4X), the predicted posterior boundary of the anterior Hb domain shifts by -8% or +9.3%, which is very close to the experimental measured value of -6.5% or +9.4%[13], rather than ±11.6% as predicted by a simple threshold activation model[29].

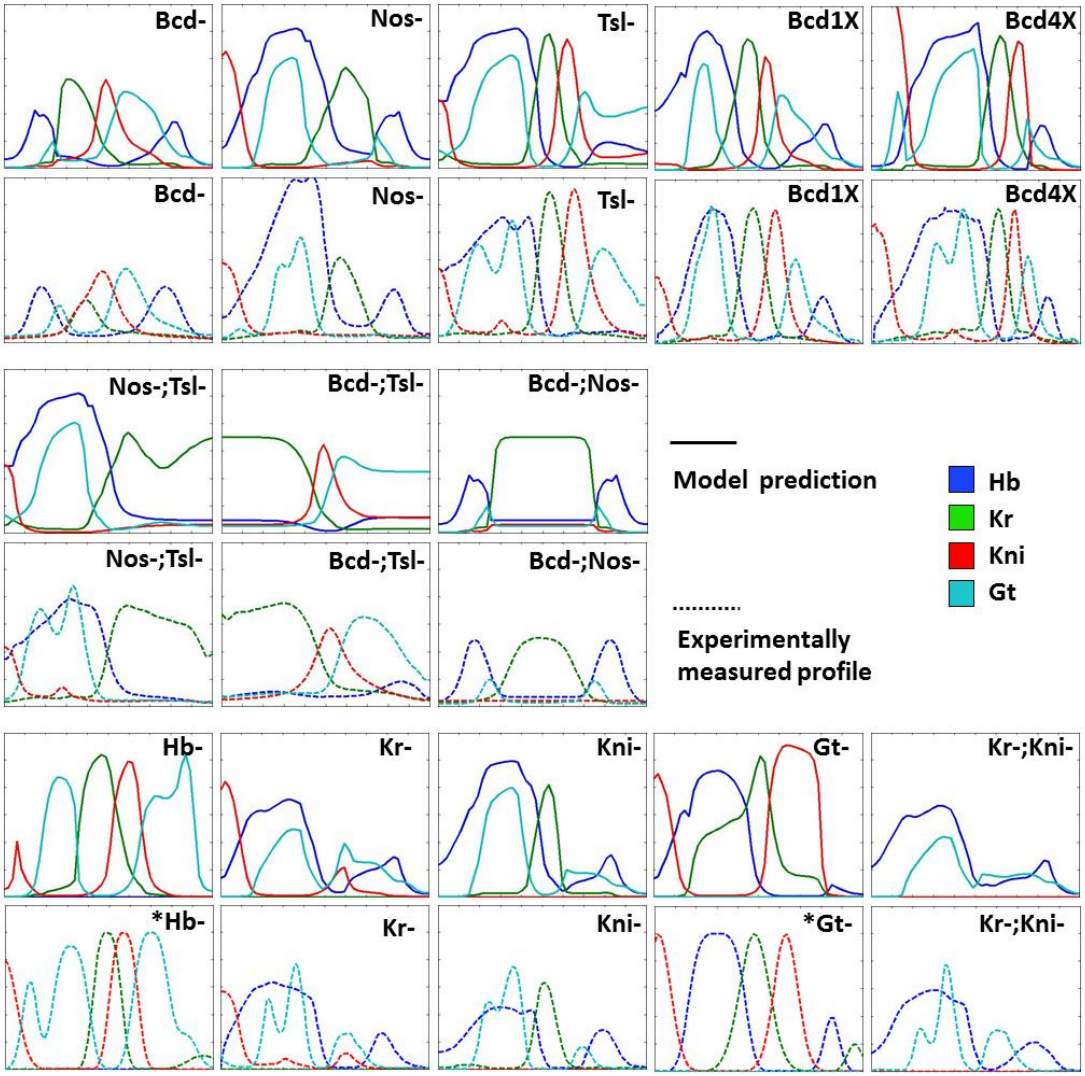

Figure 3. Comparison of the model prediction with experimental data. The predicted expression profiles (solid lines, odd rows) of Hb, Kr, Kni and Gt along the A-P axis are consistent with the experimental data (dished lines, even rows)[13,28,30-32]. For Hb^- and Gt^- data (marked with asterisks), only the numbers and rough positions of peaks are for comparison due to the semi-quantitative nature of the data.

To make a more quantitative comparison, we mark the positions of important features in the expression profiles, i.e. the main peaks and their boundaries, and then compare these positions between model prediction and the experiments (Fig. 4A, see Supplementary Information S5 for detailed algorithm). The trained DNN model shows excellent performance: nearly 90% of feature points are matched (Fig. 4B inset) and those matched features have similar experiment and predicted positions (main

scatter plot in Fig. 4B). We trained the model eight times independently; matched feature points of the resulting predictions are usually between 80% and 90% (See Supplementary Information S6 for detailed statistics).

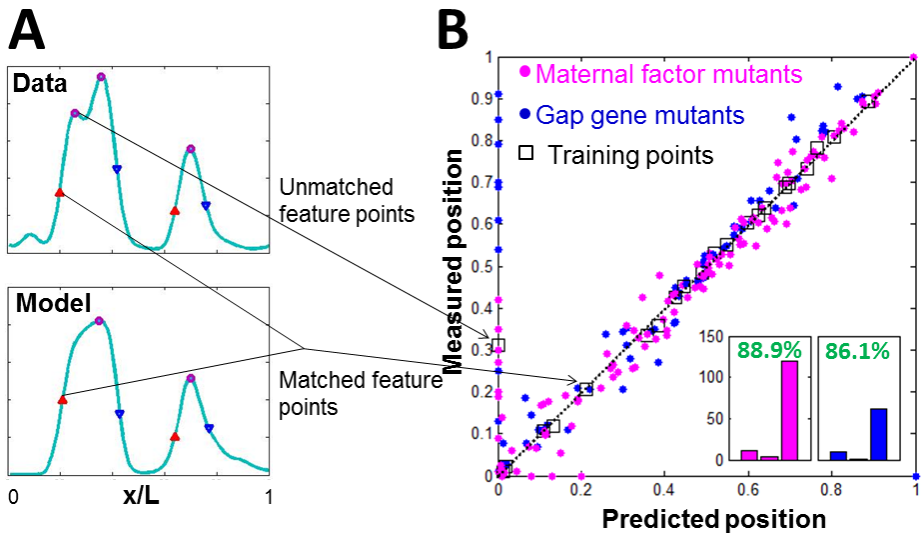

Figure 4. Quantitative visualization of the expression profile features between the DNN model and experiments. (A) Peaks and boundaries are extracted and matched between model prediction and the experiment profiles. (B) Perfect match between the predicted and the actual position would fall on the diagonal. The results for maternal factor (magenta) and gap gene (blue) mutants along with the wt data used for training (black square) are plotted as a scatter graph. The feature points that exist in data but not in prediction (or vice versa) are plotted on vertical (or horizontal) axis. Inset: For maternal factor mutants (magenta) and gap gene mutants (blue), the numbers of unmatched feature points on the vertical and horizontal axis are plotted as the first and second bar from the left. Respectively, the number of matched feature-point pairs is represented as the third bar, and its percentage is written above the bar. Nearly 90% of feature points are correctly predicted, and the predicted positions are also consistent with where they actually are.

Furthermore, the model could have some other predictions on multi-mutants that also agree with published experiments (Fig. 5). Especially, it has been reported that Nos$^-$, a severe mutant lacking almost all abdomen patterns, could be rescued by knocking out

maternal Hb (mHb) completely[33], and the above trained DNN model predicted that the gap gene profiles is very similar to wt if initial Hb is absolutely zero in a Nos- background (Fig. 5A). This result holds across eight different trainings (Fig. S6B).

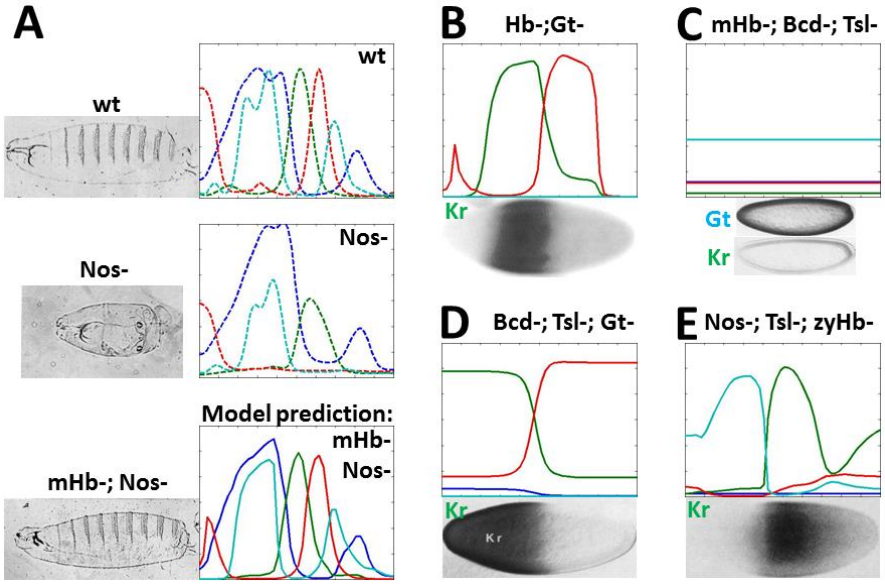

Figure 5. Further validation with some non-quantitative experimental evidence. (A) Nos$^-$ mutant differs significantly from wt both in cuticle samples (representation of larva morphology) and gap gene profiles. However this mutant can be dramatically rescued by further knocking out maternal Hb (mHb), resulting in a normal cuticle phenotype[33]. Our model successfully predicts normal gap gene profiles in this double mutant. Some other predictions consistent with experimental evidences include: (B) double mutant of Hb$^-$;Gt$^-$[34], (C) tipple mutant of mHb$^-$;Bcd$^-$;Tsl$^-$[35], (D) Triple mutant Gt$^-$;Bcd$^-$;Tsl$^-$[35], and (E) zygotic Hb mutation in Nos$^-$;Tsl$^-$ embryos[35].

Other examples of good prediction include: Kr peak still exists and expands toward anterior when Hb and Gt were knocked out simultaneously (Fig. 5B)[34]; Gt instead of Kr has uniform high expression if mHb is further knocked out in the maternal morphogen mutant Bcd$^-$;Tsl$^-$ (Fig. 5C)[35]; In Bcd$^-$;Tsl$^-$ embryos, Kr pattern remains almost the same even if Gt, which is usually thought to strongly repress Kr, is knocked out (Fig. 5D)[35]; In Nos$^-$;Tsl$^-$ embryos, mutation in zygotic Hb (mHb unaffected) will shift the anterior boundary of Kr from 50% to about 40% (Fig. 5E)[35] .

Regulation Network

Excellent prediction on nontrivial experimental observations suggests that the trained DNN model might have faithfully captured the essential characteristics of the fly embryos' developmental system. Thus decoding the black box of the DNN model should help us understanding the underlying mechanism. Here, the black box is a function calculating four output synthesis rates from seven input concentrations. Decoding stands for regenerating this input-output relation, at least partially, with a simpler and more understandable function form. As preliminary trails, we tried to extract a simple gap gene regulation network from the DNN model, and compared it with previous knowledge.

We have tried various different methods to map a deep neural network into a simple regulatory network (Fig. 6), for example, by measuring outputs of one-hot inputs (leaving one input as 1 and setting the rest to 0), calculating correlation functions between input dimensions and output dimensions, trying to fit the black box with a linear model,

$$\boldsymbol{F}_{DNN}(\boldsymbol{g}) \approx W * \boldsymbol{g}, \tag{3}$$

or a single layer neural network with shared bias values,

$$\boldsymbol{F}_{DNN}(\boldsymbol{g}) \approx sigmoid(W * \boldsymbol{g} + b). \tag{4}$$

It is unsurprising that each method, with limited plasticity, captures different aspect of the nonlinear black box, resulting in different network topologies. This result seems to undermine the legitimacy of representing such a complex system with just a simple regulation network. Though on the other hand, the extracted gap gene regulation network is qualitatively compatible with the known one deduced from experimental evidence according to reviewing literature (Fig. 6E). Albeit such similarity with existing knowledge, it is impossible to regenerate gap gene patterns from these fitted

regulation rules, suggesting that these representations are probably already over simplified. There should be some "high order effects"[36] that cannot be ignored.

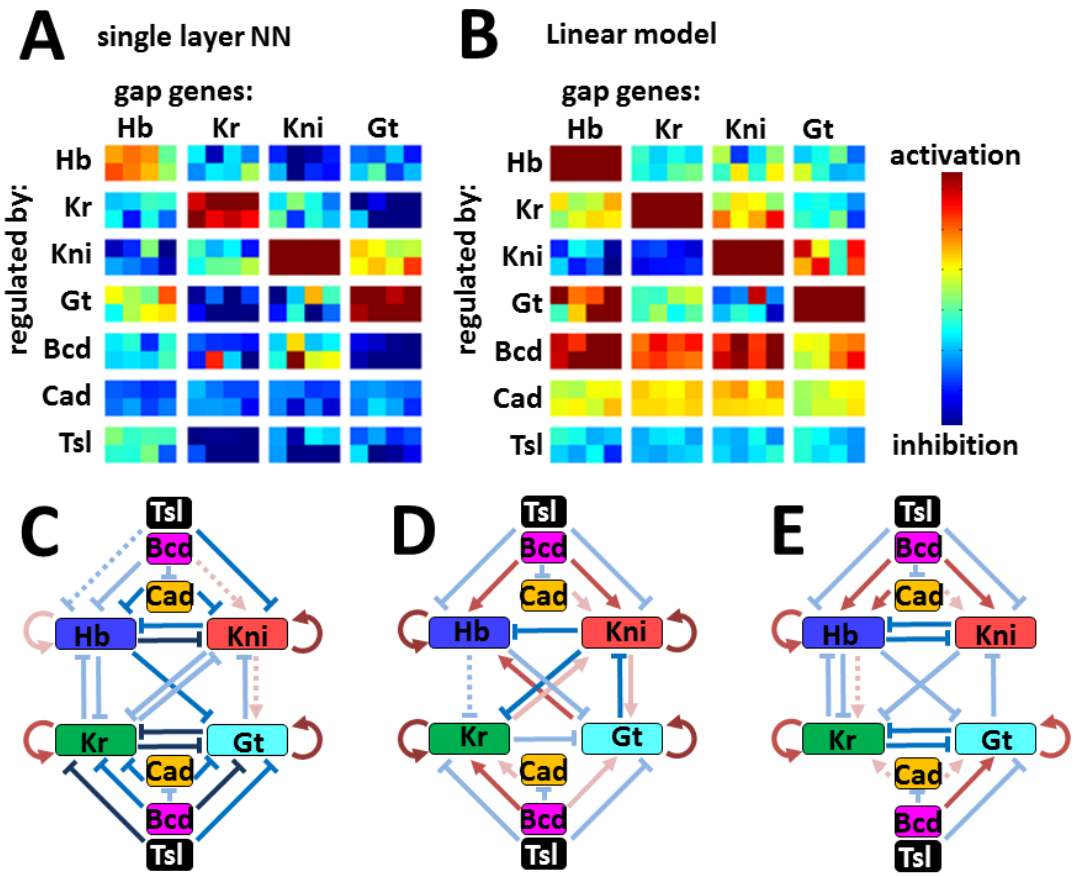

Figure 6. Visualization of the gap gene regulation networks exacted from DNN. (A) Fitting the input-output relationship of the DNN with a single layer neural network with shared bias values and sigmoid activation functions, and (B) liner model without bias. Each block of cells stands for the regulation strength on the gene name above it by the gene name left to it. And the 8 colored pixels in each block stand for eight independent training trials. The strength of the activation (repression) is colored as red (blue). The average regulation is further visualized in (C-D), with the red arrows and blue lines denote activation and repression, respectively. Darker colors represent greater regulation strength, and dash lines represent uncertain regulations, which is too weak or seriously inconsistent among different trials. (E) Regulations deduced from experimental evidence[1].

At least part of the problems in mapping the DNN to a simple regulatory network can be understood by the "inherent plasticity". For example, it is commonly accepted that Cad and its repressor Bcd both activate Hb, forming an incoherent feed forward (IFF) motif[37,38] (Fig. 6E). But as shown in Fig. 6C, an IFF motif emerges with both Bcd and Cad inhibit Hb. As Cad is set to be fully determined by Bcd in our model, these two ways of implementing an IFF motif could be functionally indistinguishable, unless Cad^- mutant is introduced. Apart from this explicit 3-node example, such degeneracy (different regulatory structure, almost identical function) could exist in a more dispersive and obscure manner on a larger network scale, making reverse engineering a unique regulation network purely from limited amount of data difficult.

However, it should be noted that for simple problems, such as "how can two-node reaction diffusion system generate stripes", this training and decoding methodology works pretty well and yields the Turing pattern mechanism (See Supplementary Information S7 for details). So whether such decoding would yield meaningful mechanism is obviously case and data dependent.

Higher Order Effects

Though it was evident above that there should be some irreducible higher order effect for gap gene pattern formation, visualization of some intersections of the high dimensional F shows smooth regulation functions, instead of a rugged landscape in typical overfitting cases. Also, among all the variations in output F when generating the wt and mutant patterns, 86.5% could be explained by a linear model (measured with Euclidean distance); i.e., after regulation matrix W being fitted:

$$\frac{\text{remianing error}}{\text{total varience}} \equiv \frac{|\boldsymbol{F}_{DNN}(\boldsymbol{g}) - W * \boldsymbol{g}|^2}{|\boldsymbol{F}_{DNN}(\boldsymbol{g})|^2} \approx 13.5\% \qquad (5)$$

Distribution of the remaining errors can also be plotted as a histogram (Fig. 7A). In most situations (for most input $\boldsymbol{g}$), error of the linear fitting is rather small,

corresponding to simple monotonic regulatory logics as the cases shown Fig. 7B. The distribution of each gene component of those inputs when linear fittings have large errors is plotted as Fig. 7A inset. These histograms show roughly where the higher order effects are. For example, Tsl seems not to be involved in those high order effects: Tsl level is low (peaked at around 0) in all those situations where linear fitting fails, thus the regulation function is almost linear when Tsl is high, which is sufficient but not necessary for concluding that Tsl effect is almost additive.

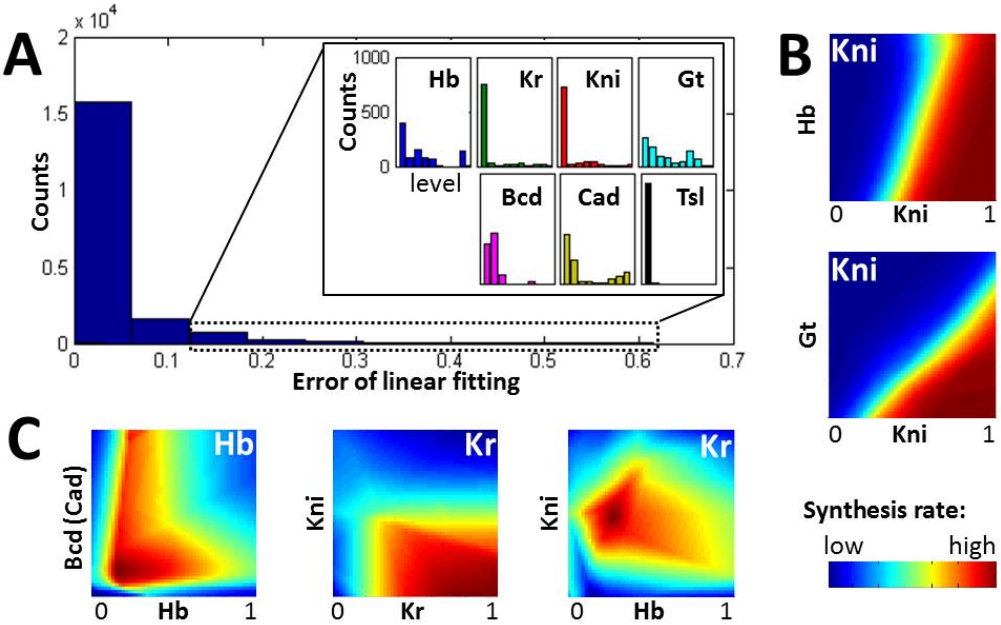

Figure 7. High order effects in regulation. (A) Histogram of the errors of linear fitting. In most situations errors are rather small, implying that simple additive regulatory logics capture a significant part of the regulation. Inset: expression level distribution among those inputs where linear fitting fails. (B) Examples of "simple" regulations, the synthesis rate of Kni is plotted as function of Hb (or Gt) and the level Kni itself, with all other factors fixed. (For the first case, Kr=0.9, Gt=0, Bcd=0.05, Tsl=0, corresponding to where the peaks of Hb and Kni overlap in wt pattern. For the second, Hb=0, Kr=0, Bcd=0.03, Tsl=0. Cad is always determined by Bcd using Eq. S1). Inhibition by Hb (or Gt) and Kni self-activation function almost additively. (C) Examples of some high order effects, noting the non-monotonic self-regulation of Hb (Kr, Kni, Gt, Tsl=0).

On the other hand, both Hb and Kr show some high order effects (Fig. 7C). First, Hb is self-activating at low levels, but self-inhibiting at high levels, when Bcd>0.1, corresponding to 0% to 38% embryo length. Interestingly, evidences for both self-activating[39,40] and self-inhibiting[15,41] are reported previously. Second, self-activation of Kr and the inhibition from Kni on Kr acts like an AND-gate (Keeping Hb=0, Gt=0, Bcd=0.05, Tsl=0). Thirdly, both Hb and Kni activate Kr at low concentrations but inhibit at higher concentrations (Keeping Kr=0.1, Gt=0, Bcd=0.05, Tsl=0). This is in accordance with the reported dual regulation effect (both activation and inhibition) of Hb on Kr[42]. Notably, only the observation of the self-regulation of Hb, not the last two, is consistent among different training trails (Fig. S6B).

Discussion

No Better Prediction with More Training Data

Intuitively, with more training data added, model parameters should be more tightly constrained, resulting in better solutions; however, this is not the case for the current situation. If extra snapshots of three maternal morphogen mutants (Bcd^-, Nos^- and Tsl^-) were added to the training set, predictions are significantly interfered. Correctly predicted features drop from 88.9% and 86.1% to 82.6% and 72.6% for maternal factor mutants and gap gene mutants, respectively (Fig. 8A). Similar results in prediction reappear if we add other extra mutant profiles to the training set. We suspect that overfitting might be caused by some potential incompatibility within the data set, as DNN is capable of fitting all sorts of features. Such incompatibility may be reduced by more careful background removal, expression level normalization, embryo age estimation, etc. For comparison, if trained with wt only, the model yields even better predictions on gap gene mutants (89.2% *vs.* the original 86.1%, Fig. 8B), but much worse on maternal factor mutants. Again, these results reflect some inherent plasticity: the model seems to be able to correctly predict gap gene mutants even

without a correct "understanding" of the role of each upstream morphogens.

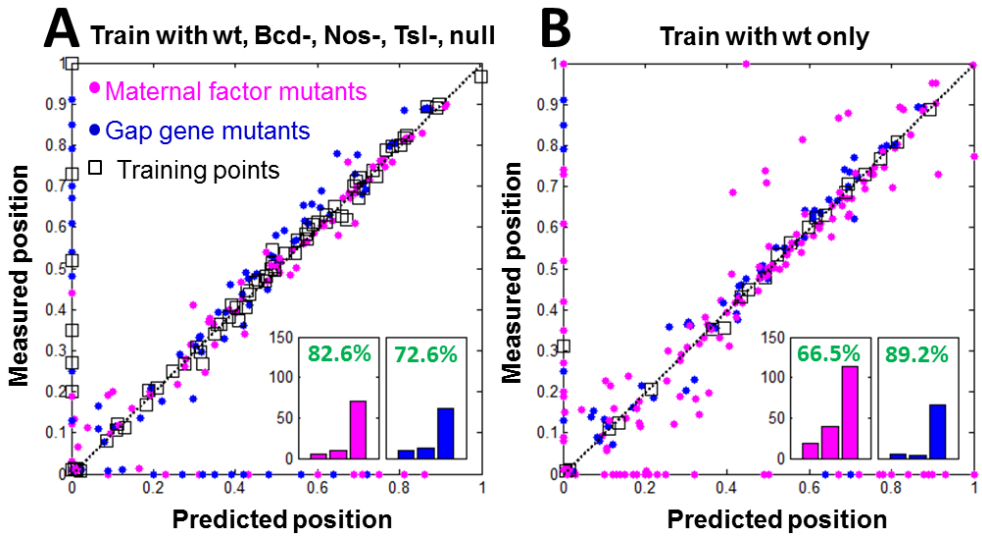

Figure 8. Profile features between the experiments and the prediction of the DNN model with different training sets. Main peaks and boundary positions of gap gene expression profiles based on the DNN model trained with seven-frame wt time series, snap shots of the three maternal single mutant Bcd$^-$, Nos$^-$, Tsl$^-$ and maternal triple mutant Bcd$^-$;Nos$^-$;Tsl$^-$ at about 40 min in nc14 (A); and with seven-frame wt time series only (B). Marks are the same as Figure 4.

Robustness against Missing Factors

It can never be guaranteed in practice that no factors (hidden genes, gene modification, small RNA, etc.) are left unknown. Instead of wishing the missing factors are not important, we can demonstrate that our model is insensitive to missing even important factors. With Kni pretended to be absolutely unknown, i.e. removed from data and model, we retrained the three-node model, and remarkably it still yielded good predictions on features of the remaining gap genes (Fig. 9A).

The regulation network reconstructed by the method discussed in Fig. 6, though rough, bears some hint for how Kni's role was effectively absorbed by other genes (Fig. 9B). For example, the original (Fig. 6C) double inhibition (Hb inhibits Kni and Kni

inhibits Kr) is replaced by an effective activation from Hb to Kr.

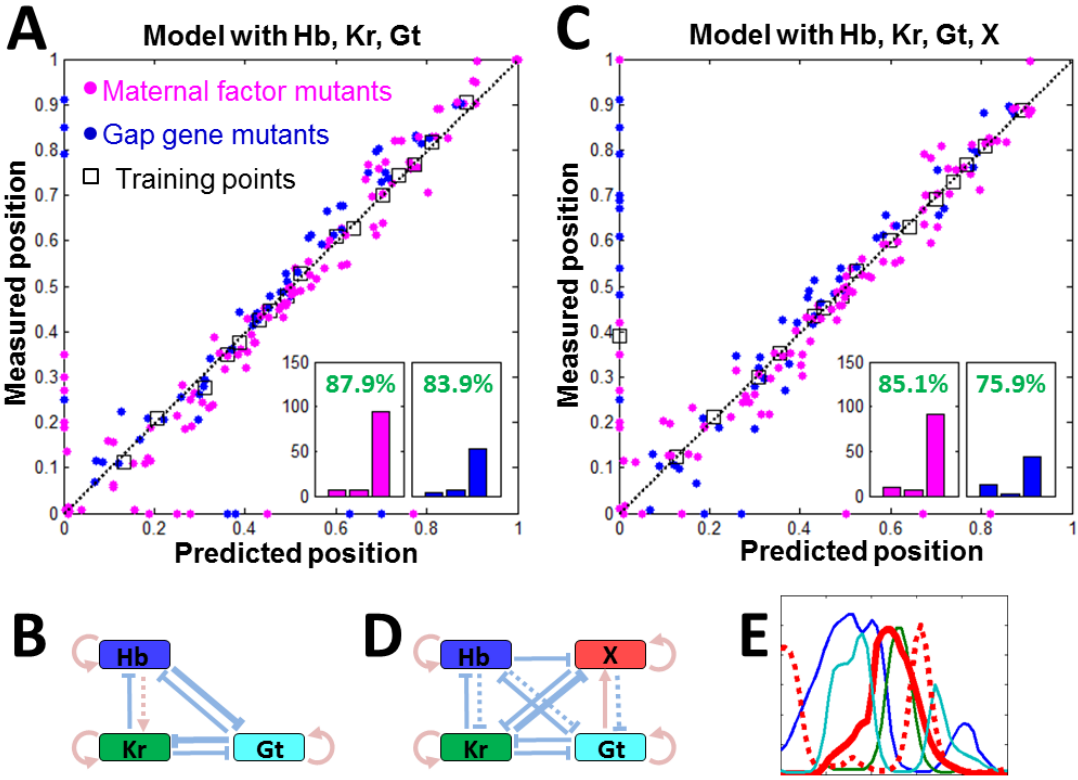

Figure 9. Missing important genes will not make our model collapse. (A) Excluding Kni, simulating profiles of the remaining three gap genes with a three-node model, prediction performance only drops a little (~1%), suggesting that the model is not sensitive to missing genes. (B) Corresponding regulation network with Kni being "absorbed" (roles of maternal factors unchanged, not shown). (C) An extra free node X added to the three-node model does not help with mutant prediction (even worsen), and its regulation (D) or pattern (E, red solid line) did not take up the role of real Kni (red dotted line in E).

This result may serve as a demonstration to how the model can work robustly with missing factors. However, such robustness may hinder the model's ability to discover new genes. Ideally if an extra node is provided to help pattern formation freely, while an irreplaceable factor is missing, this free node would be able to take the role of the missing one. In simple cases like the three-node adaptation network, the buffering node automatically emerges if trained in this way (Supplementary S7). But it is not

the case for more complex situations: here, an additional free node X did not help with better prediction (Fig. 9C), nor did it show the pattern or regulation of the original Kni (Fig. 9D-E).

It should be noted that overall introducing genes with known patterns usually help with prediction performance. As a good example, Cad helps significantly improving predictions, though theoretically effects of Cad can always be absorbed as a nonlinearity of Bcd regulation function.

Alternative Mechanism

With previous models, it has been difficult to explain the global decline of gap gene profiles after 40 minutes in nc14 without any change in external inputs[6,8]. It has been suggested that this phenomenon could be attributed to the events associated with maternal-zygotic transition, such as the decaying of the Bcd gradients in nc14[43], the turn off of the Bcd transcription regulation on Hb[44], or the switch of the Hb enhancer[15]. While we could capture the falling phase of the gap gene profiles if we introduce the shutdown of Bcd in our model in early nc14, surprisingly, we can also train a model with both the rising phase before 40 minutes and the falling phase from 40 to 58 minutes without any input change. The resulting model can not only fit the decline phase well, but also have reasonably good predictions on mutants' profiles (78.3% feature points in maternal factor mutants, and 85.1% in gap gene mutants were correctly predicted).

In the same sense, our present model did not take into account of many factors, such as diffusion, lifetime of mRNAs, the time delay due to transcription and translation and degradation of the maternal morphogens, but it still has satisfactory predictions, suggesting that those effects are not irreplaceable for the formation of the main pattern structures.

Conclusion

Differential equation models have been widely and successfully used in simulating and understanding biological systems. However, it is evident that such models, in its conventional implementation, can often run into their limitations in dealing with systems of high complexity. Part of the problem may come from the standard modeling procedure: (1) (qualitative) regulation relations are extracted/inferred from experimental observation/data, which are typically obtained by perturbing the system in a few limited and mostly qualitative ways (e.g. deleting, mutating and overexpressing genes of interest); and (2) predefined simple functional forms (e.g. Hill functions) are used to model the regulations with some parameters. Information can get lost in both steps, and the resulting model can be too restricted and confined to reflect the true essential dynamics of the system. Therefore, it may be worthwhile to try to use the available data differently. The approach we adopted here with DNN takes the data in its entirety – the expression profiles of the gap genes. The fact that our model can acquire such an impressive predictive power with only the wt dynamics data is also suggestive – there is a rich content of information in the dynamics of the system as compared with the end phenotype.

Albeit the difficult interpretability like all DNN models, our model did generate some new insight about the patterning system in early fly embryos. More importantly, with such an *in-silico* model one can conceivably perform almost arbitrary perturbations and thought experiments, which would otherwise be difficult to perform in wet experiments and be less reliable in the conventional network models. So maybe in the near future, this approach could become a powerful lens to provide novel insights and new perspectives, contributing to our understanding of complex systems.

Acknowledgements

We thank Xiaojing Yang, Ning Yang, and Xiao Li for helpful discussions. The work was supported by the Chinese Ministry of Science and Technology (Grant No.2015CB910300) and the National Natural Science Foundation of China (Grant No. 91430217).

References

1. Jaeger, J. The gap gene network. *CMLS, Cell. Mol. Life Sci.* **68,** 243–274 (2010).
2. Houchmandzadeh, B., Wieschaus, E. & Leibler, S. Establishment of developmental precision and proportions in the early Drosophila embryo. *Nature* **415,** 798–802 (2002).
3. Gregor, T., Tank, D. W., Wieschaus, E. F. & Bialek, W. Probing the Limits to Positional Information. *Cell* **130,** 153–164 (2007).
4. Gregor, T., Wieschaus, E. F., McGregor, A. P., Bialek, W. & Tank, D. W. Stability and Nuclear Dynamics of the Bicoid Morphogen Gradient. *Cell* **130,** 141–152 (2007).
5. Dubuis, J. O., Samanta, R. & Gregor, T. Accurate measurements of dynamics and reproducibility in small genetic networks. *Molecular Systems Biology* **9,** 1–15 (2013).
6. Jaeger, J. *et al.* Dynamic control of positional information in the early Drosophila embryo. *Nature* **430,** 368–371 (2004).
7. Manu *et al.* Canalization of Gene Expression in the Drosophila Blastoderm by Gap Gene Cross Regulation. *PLoS Biol* **7,** e49 (2009).
8. Papatsenko, D. & Levine, M. The Drosophila gap gene network is composed of two parallel toggle switches. *PLoS ONE* **6,** e21145 (2011).
9. Segal, E., Raveh-Sadka, T., Schroeder, M., Unnerstall, U. & Gaul, U. Predicting expression patterns from regulatory sequence in Drosophila segmentation. *Nature* **451,** 535–540 (2008).
10. Huang, A., Amourda, C., Zhang, S., Tolwinski, N. S. & Saunders, T. E. Decoding temporal interpretation of the morphogen Bicoid in the early Drosophila embryo. *eLife* **6,** e26258 (2017).
11. Bothma, J. P., Norstad, M. R., Alamos, S. & Garcia, H. G. LlamaTags: A Versatile Tool to Image Transcription Factor Dynamics in Live Embryos. *Cell*

173, 1810–1822 (2018).
12. Desponds, J., Tran, H., Ferraro, T. *et al.* Precision of Readout at the *hunchback* Gene: Analyzing Short Transcription Time Traces in Living Fly Embryos. *PLoS Comput Biol* **12,** e1005256.d (2016).
13. Liu, F., Morrison, A. H. & Gregor, T. Dynamic interpretation of maternal inputs by the Drosophila segmentation gene network. *Proceedings of the National Academy of Sciences* **110,** 6724–6729 (2013).
14. Schroeder, M. D. *et al.* Transcriptional Control in the Segmentation Gene Network of Drosophila. *PLoS Biol* **2,** e271 (2004).
15. Perry, M. W., Bothma, J. P., Luu, R. D. & Levine, M. Precision of Hunchback Expression in the Drosophila Embryo. *Current Biology* **22,** 2247–2252 (2012).
16. El-Sherif, E. & Levine, M. Shadow Enhancers Mediate Dynamic Shifts of Gap Gene Expression in the Drosophila Embryo. *Current Biology* **26,** 1164–1169 (2016).
17. Li, X.-Y. *et al.* Transcription Factors Bind Thousands of Active and Inactive Regions in the Drosophila Blastoderm. *PLoS Biol* **6,** e27 (2008).
18. Estrada, J., Wong, F., DePace, A. & Gunawardena, J. Information Integration and Energy Expenditure in Gene Regulation. *Cell* **166,** 234–244 (2016).
19. Daniels, B. C. & Nemenman, I. Automated adaptive inference of phenomenological dynamical models. *Nature Communications* **6,** 1–8 (2015).
20. Camacho, D. M., Collins, K. M., Powers, R. K., Costello, J. C. & Collins, J. J. Next-Generation Machine Learning for Biological Networks. *Cell* **173,** 1581–1592 (2018).
21. Angermueller, C., Pärnamaa, T., Parts, L. & Stegle, O. Deep learning for computational biology. *Molecular Systems Biology* **12,** 878 (2016).
22. Ching, T. *et al.* Opportunities and obstacles for deep learning in biology and medicine. *Journal of The Royal Society Interface* **15,** 20170387 (2018).
23. Poggio, T., Kawaguchi, K., Liao, Q., Miranda, B. & Rosasco, L. Theory of Deep Learning III: explaining thenon-overfitting puzzle. 1–36 (2018).
24. Jaeger, J. & Reinitz, J. On the dynamic nature of positional information. *Bioessays* **28,** 1102–1111 (2006).
25. Surkova, S. *et al.* Characterization of the Drosophila segment determination morphome. *Developmental Biology* **313,** 844–862 (2008).
26. Wreden, C., Verrotti, A. C., Schisa, J. A. & Lieberfarb, M. E. *Nanos* and *pumilio* establish embryonic polarity in Drosophila by promoting posterior deadenylation of hunchback mRNA. *Development* **124,** 3015-3023 (1997).
27. Wang, C. & Lehmann, R. Nanos Is the Localized Posterior Determinant in Drosophila. *Cell* **66,** 637–647 (1991).
28. Petkova, M. D., Tkačik, G. & Gregor, T. Optimal decoding of informaiton from a genetic network. 1–19 (2016).
29. Wolpert, L. Positional information and patterning revisited. *Journal of Theoretical Biology* **269,** 359–365 (2011).
30. Liu, F. & Gruebele, M. Downhill dynamics and the molecular rate of protein folding. *Chemical Physics Letters* **461,** 1–8 (2008).

31. Kozlov, K., Surkova, S., Myasnikova, E., Reinitz, J. & Samsonova, M. Modeling of Gap Gene Expression in Drosophila Kruppel Mutants. *PLoS Comput Biol* **8,** e1002635 (2012).
32. Surkova, S. *et al.* Developmental Biology. *Developmental Biology* **376,** 99–112 (2013).
33. Irish, V., Lehmann, R. & Akam, M. The Drosophila posterior-group gene nanos functions by repressing hunchback activity. *Nature* **338,** 646–648 (1989).
34. Kraut, R. & Levine, M. Mutually repressive interactions between the gap genes giant and Krüppel define middle body regions of the Drosophila embryo. *Development* **111,** 611–621 (1991).
35. Struhl, G., Johnston, P. & Lawrence, P. A. Control of Drosophila body pattern by the hunchback morphogen gradient. *Cell* **69,** 237-249 (1992).
36. Basu, S., Kumbier, K., Brown, J. B. & Yu, B. Iterative random forests to discover predictive and stable high-order interactions. *Proceedings of the National Academy of Sciences* **115,** 1943–1948 (2018).
37. Mangan, S. & Alon, U. Structure and function of the feed-forward loop network motif. *Proc. Natl. Acad. Sci. U.S.A.* **100,** 11980–11985 (2003).
38. Alon, U. Network motifs: theory and experimental approaches. *Nature Publishing Group* **8,** 450–461 (2007).
39. Treisman, J. & Desplan, C. The Products of the Drosophila Gap Genes Hunchback and Kruppel Bind to the Hunchback Promoters. *Nature* **341,** 335–337 (1989).
40. Lopes, F. J. P., Vieira, F. M. C., Holloway, D. M., Bisch, P. M. & Spirov, A. V. Spatial Bistability Generates hunchback Expression Sharpness in the Drosophila Embryo. *PLoS Comput Biol* **4,** e1000184 (2008).
41. Xu, H., Sepúlveda, L. A., Figard, L., Sokac, A. M. & Golding, I. Combining protein and mRNA quantification to decipher transcriptional regulation. *Nature Methods* **12,** 739–742 (2015).
42. Holloway, D. M. & Spirov, A. V. Mid-Embryo Patterning and Precision in Drosophila Segmentation: Krüppel Dual Regulation of hunchback. *PLoS ONE* **10,** e0118450 (2015).
43. Verd, B., Crombach, A. & Jaeger, J. Dynamic Maternal Gradients Control Timing and Shift-Rates for Drosophila Gap Gene Expression. *PLoS Comput Biol* **13,** e1005285 (2017).
44. Liu, J., Xiao, Y., Zhang, T. & Ma, J. Time to move on: Modeling transcription dynamics during an embryonic transition away from maternal control. *Fly* **10,** 101–107 (2016).

Supplementary Information for

Toward deciphering developmental patterning with deep neural network

Jingxiang Shen, Feng Liu, Chao Tang

S1. Maternal Inputs

Our model takes into account three classes of maternal morphogens, anterior Bicoid (Bcd), posterior Nanos (Nos), and terminal Torso-like (Tsl) as representations[1-3]. Steady-state profiles of Bcd and Tsl are used as direct inputs. The Bcd profile comes from direct quantitative measurement[4], whereas the Tsl profile is deduced from that of Capicua, which is supposed to be inhibited by Torso pathway[5]. Nos is introduced implicitly as the profile of initial Hb (denoted as mHb), because the main function of Nos is the translational repression of maternal Hb thus shaping the initial condition of Hb[6] (Fig. S1A).

Usually other two downstream morphogens Cad and Tll are also considered as extra inputs in previous models[7-9], though they might not bring in new positional information. Cad is vital for abdominal pattern[10], and binds strongly to the cis-regulation sequences of many gap genes[11], and is also included it in our model. Pattern of Cad is thought to be mainly shaped by translational repression from Bcd[12], so we fit its pattern[13] as a function of Bcd in wt, and assume this relation holds in all mutants:

$$Cad = 0.0008/(0.0008 + Bcd^{2.5}) \qquad (S.1)$$

As for Tll, known mutants suggests that it is regulated in a much more complex manner, influenced nonlinearly at least by Bcd and Tsl[14], thus we could not fit it with a simple function. So the effect of Tll is left to be absorbed by DNN, considering the pattern of Tll is presumably fully dictated by the included maternal factors and nearly unchanged within the relevant time period[13].

To sum up, steady-state profiles of Bcd, Tsl and Cad selected directly as inputs. The effect of Nos is expressed as an initial pattern of Hb (mHb), while the effect of Tll is absorbed in the black box of DNN (Fig. S1A). Since Bcd and Tsl are nearly completely independent, their profiles are unchanged when other genes are mutated. In contrast, mHb and Cad are constant of their maximum expression in wt along the whole AP axis in the maternal factor mutant Nos$^-$ and Bcd$^-$, respectively (Fig. S1B).

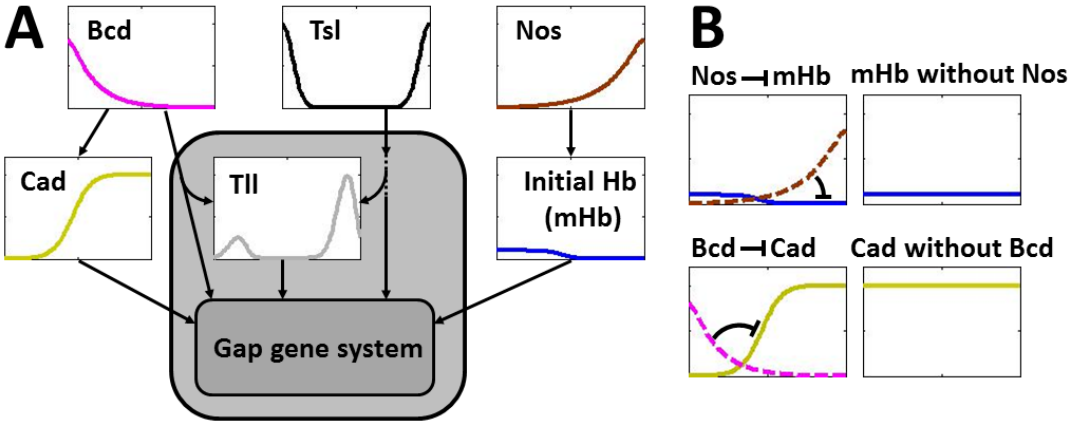

Figure S1. Maternal morphogens as upstream inputs for gap gene expression. (A) Primary positional information is thought to be carried by three factors: Bcd, Nos and Tsl. Other downstream morphogens also play roles in guiding gap gene expression but are themselves regulated by the above three factors, e.g., Cad is repressed by Bcd, mHb is repressed by Nos, and Tll is regulated by both Tsl and Bcd. These "downstream" morphogens could theoretically be absorbed into the DNN. (B) The profile of mHb in Nos$^-$ mutant is assumed to be a flat line in the whole embryo with the estimated maximum expression level of the mHb in wt. The profile of Cad in Bcd$^-$ mutant is assumed to be a flat line in the whole embryo, with the estimated maximum expression level of the Cad in wt.

S2. Gap-gene Dataset

For gap gene expression profiles, we chose the published data with the highest quality: the time course data in nc14 of wt from Gregor lab[15], a single snapshot at around 40 minute into nc14 of the maternal factor mutant from Gregor lab[16] and the gap gene mutant (Kr$^-$, Kni$^-$, and Kr$^-$;Kni$^-$) from Reinitz lab[17], all these data were obtained via

immunostaining on fixed embryos.

Since the data obtained with different experiment methods, from different labs, or even different batches may bear large systematic deviations from each other. We further processed the collected data to keep them self-consistent. To be specific: (1) we only use mean expression profile of many embryos at the same temporal stage, as subtle subjects as noise attenuation is beyond the scope of this work. (2) Expression level of the four gap genes are normalized according to their spatial maximal in wt pattern at 40 minutes nc 14, just as reference[16] (except for Hb^- & Gt^-). (3) A spatial-temporal Gaussian filter is applied to smooth the time course data. Standard deviation of the Gaussian kernel is set to be 2% embryo length spatially and 5 minutes temporally. (4) While most profiles are extracted from the dorsal side of the mid-sagittal plane of unflattened embryos, profiles of Kr^-, Kni^-, and $Kr^-;Kni^-$ double mutant were obtained from the middle surface of slightly flattened embryos. By comparing wt pattern from these two methods, a nonlinear transformation in position x should be applied to makes these three profiles have consistent coordinate system with the others.

$$(x' - 0.5) = 2.78\,(x - 0.5)^5 - 0.1364\,(x - 0.5)^4 - 1.734\,(x - 0.5)^3 + 0.07932(x - 0.5)^2 + 1.26\,(x - 0.5) - 0.0134$$

(5) Peak heights measured in flattened embryos (Kr^-, Kni^-, and $Kr^-;Kni^-$) are also adjusted so as to maintain correct ratio with corresponding peaks in wt profile measured with the same method. (6) For Hb^- and Gt^-, profiles are extracted from stained embryo images from published papers[18-21], both time point and embryo orientation are not carefully controlled, and no normalized factor is available (simply normalized with their own peak values); so only number of expression peaks and their rough positions make sense.

Note that the time course data of wt is from 8 to 58 minutes in nc14. As the first

embryo in the dataset is already at 8 min in nc14, any extrapolation of the data is purely artificial.

S3. Network Structure and Loss Function

The fully connected recurrent block has 3 hidden layers, each with 16 nodes, and an output layer with 4 nodes serving as synthesis rate. Activation function is rectified linear unit (ReLU(x) = max(x,0)) except the output layer, which uses Sigmoid (1/(1+exp(-x))), as synthesis rate should be bounded between 0 to 1. We have tried different structures, 3 to 8 layers, 4 to 64 nodes width; except for cases that are too shallow or narrow, results are almost the same.

The embryo is divided into 48 grids along the anterior-posterior axis. With no diffusion thus no spatial coupling, the number of grids obviously does not matter. Time step in the main text is set as 1.85 min, compared with data frames every 3 steps to compute loss function. Shorter (1.1 min, 5 steps between adjacent data frames) or longer (2.8 min) time steps do not affect the result ether.

Unlike typical recurrent neural networks, we did not train our model to predict frame t+1 given frame t. Instead, we start with an initial pattern, and train our model to match every following data frames. The reason is that frame t+1 profile is actually very close to that of frame t, so even a poorly trained model can predict the next frame from the previous one pretty well. The current setting here can result in much higher training accuracy by making use of the fact that error accumulates through iterations.

S4. Training

Strictly speaking, model structure, hyper parameter, or detailed setting of the training data should be tuned totally independent from the test set, i.e. using a validation set.

It's natural to assume that good performance on the validation set should guarantee good performance on test set. However in this case, a solution with very accurate prediction on gap gene mutants but bad prediction on maternal factor mutants (Fig. 7B), suggesting different mutant profiles cannot be simply viewed as sampled from the same "distribution" (it's hard to define a distribution with only a couple of mutant profiles, and for the same reason, we do not adopt batch-normalization technique). Hence we just tuned our model to have the best prediction on the results shown in figure 3. Meanwhile, results shown in Fig. 5 are purely predictions; those evidences were gathered after model training is finished.

The Euclidean errors of the maternal factor mutant predictions decay along with the training errors (Fig. S4A) and did not rise up again, indicating no obvious overfitting. Curves of gap gene mutants are not shown here, as Hb^- and Gt^- are not properly normalized, and the rest are normalized in a complex manner (see S2); Euclidean distance is no longer a good characterization of profile similarity if errors exist in normalizations.

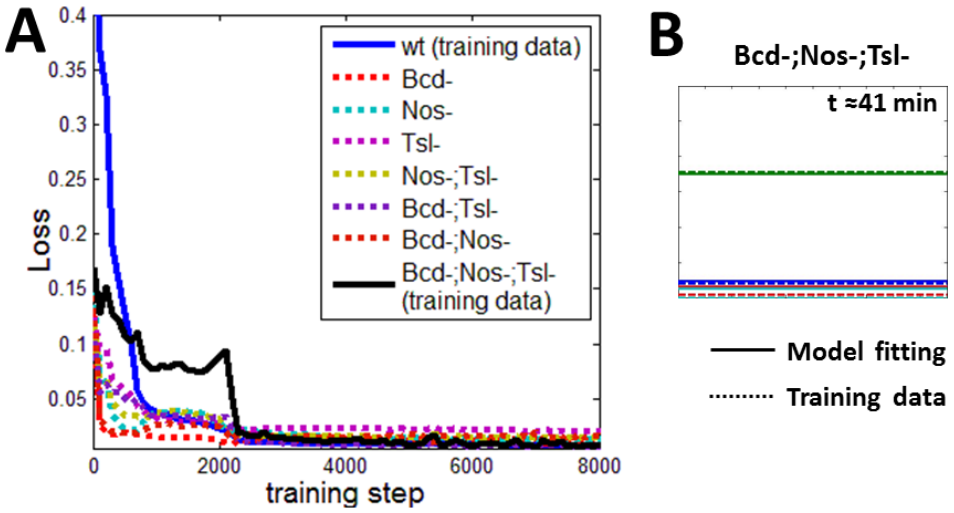

Figure S4. (A) Euclidean error of the predictions the maternal factor mutants (dotted lines) decay along with the training error (blue and black solid line), and did not rise up when further trained. (B) Fitting result is almost perfect with Bcd^-;Nos^-;Tsl^-, as supplement of Fig. 2 in the main text.

Fitting result of maternal factor triple mutant (Bcd^-;Nos^-;Tsl^-) is shown in Fig. S4B, as supplement of Fig. 2 in the main text.

S5. Matching Profile Features

Euclidean distance is a good characterization of similarity only if the profiles are differed slightly, hence can be used as loss function for training. However predictions may not necessarily be perfect at every position, as the positions of peaks and boundary of the expression profiles are often thought to carry important positional information. So we implement a simple algorithm to compare the feature points of the predicted profiles and experimentally measured profiles for quantitative evaluation of the prediction performance.

Peaks above a certain threshold (0.18) and boundaries (half height of the nearest peak) are extracted as feature points. Feature points of the same type (e.g. rising boundary of Gt) are crosschecked between predicted profiles and measured profiles. A pair is matched if point A from the predicted profile is closet to point B in the measured profiles among all feature points of the same type, and vice versa. Examples of matched and unmatched feature points are shown in Fig. 4A in main text.

S6. Independent Training Trails

We trained the model eight times independently; most results have satisfactory predictions on mutant patterns. Predictions are evaluated by the percentage of successfully matched feature points (Table S6). Fig. 2-5 in the main text and Fig. S4 in supplementary information are all results from trail No.7, marked with asterisks.

Table S6. Stable prediction performance of 8 repeated training trails.

Trails	1	2	3	4	5	6	7*	8
Maternal factor mutants	87.4%	80.3%	90.4%	90.8%	87.7%	89.5%	88.9%	89.6%
Gap gene mutants	80.3%	79.5%	70.5%	82.2%	76.0%	84.5%	86.1%	81.3%

Non-monotonic self-regulation of Hb, and predictions on Nos$^-$;mHb$^-$ mutant are also consistent across most of the training trails. (Fig. S6)

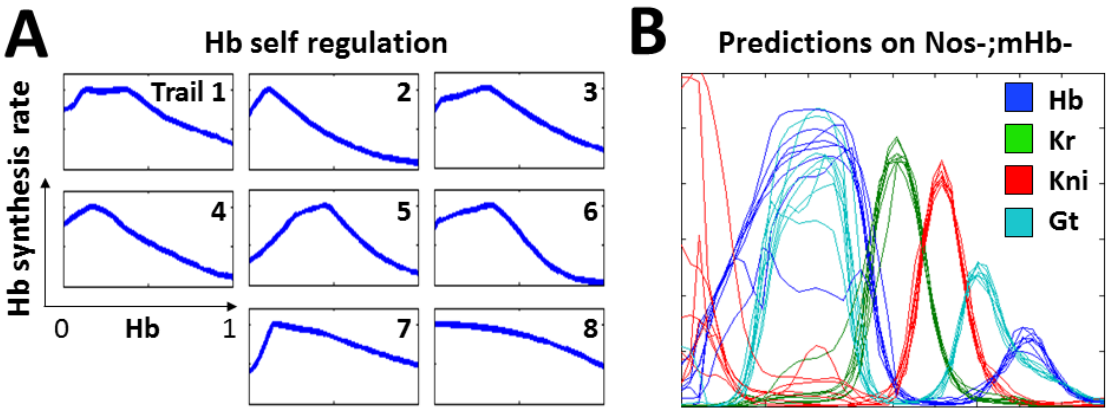

Figure S6. (A) Non-monotonic self-regulation of Hb emerges in most of the training trails. Here we plot Hb synthesis rate with Hb level vary from 0 to 1, while other gap gene level are kept at zero. Maternal morphogen levels are set corresponding to 33% embryo length in wt (Hb anterior peak). (B) The prediction that Nos$^-$ could be rescued by knocking out mHb is also conserved across these trails, especially the abdomen patterns that loosed entirely in Nos$^-$.

S7. Successful Network Reconstruction on Simpler Tasks

One major reason that regulation mechanism cannot be reliably reconstructed is the inherent placidity of the high-dimensional dynamic system itself (not the multi-layer DNN): errors can always be effectively compensated by changes of regulation in other dimensions while having little differences in overall outputs. However for

simpler tasks with lower dynamic dimensions such placidity is not overwhelming, thus reliable mechanism reconstruction is possible.

The first example is “three-node adaptation”, i.e. what kind of regulation could make output C only response to the time derivative of input A but not its absolute value, with the help of an extra node X (Fig. S7A). The system contains two variables X and C, and an upstream input A, forming a set of ordinary differential equations.

Again, we express the synthesis term as a DNN of the same architecture as discussed in S3 except one layer less. Loss function is defined as: (1) the system should be at fixed point when A=0, C=0.5 and A=1, C=0.5. (2) When A jumps from 0 to 1 (denote as t=0.1), temporal average of C should be as large as possible during $0.1<t<0.2$. (3) No constrain on the output during $0.2<t<0.3$. (4) Output C should return to 0.5 after t=0.3. Note that we did not explicitly determine the behavior of node X. The model yields perfect adaptation after training (Fig. S7B).

As the system is only two dimensional, the regulation function can be fully expressed with a vector field; and the adaptation process can be visualized explicitly (Fig. S7C). Activation and inhibition could be directly read out from the vector field, and fortunately non-monotonic regulation did not appear (Fig. S7D). Enumeration of three-node adaptation networks[22], though under constrain of monotonic regulation, yields two “adapting motifs”: incoherent feed forward (IFF) loop and feedback loop. Both motifs reappeared in our result, and free node X automatically took up the role of buffering node in IFF loop.

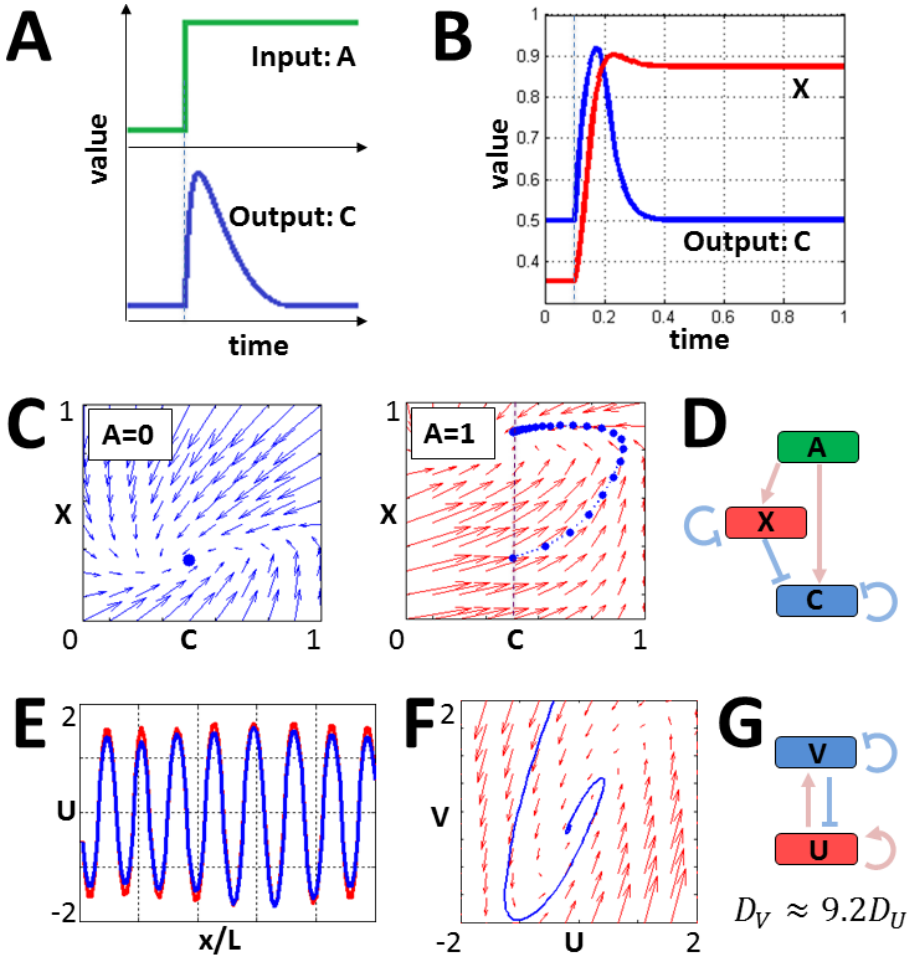

Figure S7. Revealing three-node adaptation and Turing pattern with DNN. (A-D) Problem definition, fitting result, visualization of regulation function, and reconstructed network for 3-node adaptation task. Resulting network topology is consistent with previous enumeration results[22]. (E-G) Stripe formation with 2-node reaction-diffusion dynamic. After trained, the DNN turned out to adopt Turing bifurcation mechanism[23].

A second example, "how to form stripe of a given wavelength with a two-node reaction-diffusion system", yields Turing pattern[23] mechanism. Reaction term (both synthesis and degradation together) are expressed as a DNN with the same architecture as the previous one. Diffusion is introduced as a convolution layer with Gaussian kernels, whose standard deviations (length dimension) are trainable variables (proportional to square root of diffusion constants, dimension length^2/time).

Boundary condition is set to be periodic. Loss function is defined as two parts: (1) Fourier spectrum of V should be close to 1 at k_0=0.5, and close to 0 elsewhere; (2) pattern of U should be close to V.

After training, the model shows perfect stripes with k≈k_0, regardless of domain size (Fig. S7E); though k may be distorted a little so as to maintain that domain contains integer number of wavelengths. And the underlying mechanism turned out to be Turing bifurcation: the regulation function is again visualized as a vector field, which is basically a stable spiral (Fig. 7F). Linear expansion around the fixed point yields Jacobian matrix:

$$J = \begin{pmatrix} 0.72 & 1.71 \\ -0.4 & -0.9 \end{pmatrix}$$

Linear stability can thus be checked, confirming that it is actually a stable spiral.

$$\mathrm{Tr}(J) = -0.18 < 0; \;\; \det(J) = 0.0369 > 0$$

Qualitatively, U serves as activator and V serves as an inhibitor, and diffusion constant of V is much greater than that of U (Fig. S7G).

$$D_V = 9.198\, D_U$$

Further, with Jacobian matrix and ratio between diffusion constants, it is possible to calculate the criteria for Turing bifurcation, which is satisfied:

$$0 > \det(J) - \frac{(J_{11}D_V + J_{22}D_U)^2}{4D_U D_V} = -0.85$$

And the most unstable wavenumber, which turned out to be approximately the previous assigned value:

$$k_c = \sqrt{\frac{J_{11}D_V + J_{22}D_U}{2D_U D_V}} = 0.618 \approx k_0 = 0.5$$

Reference

1. Driever, W. & Nüsslein-Volhard, The *bicoid* protein determines position in the *Drosophila* embryo in a concentration-dependent manner. *Cell* **54,** 95–104 (1988).
2. Wang, C. & Lehmann, R. Nanos Is the Localized Posterior Determinant in Drosophila. *Cell* **66,** 637–647 (1991).
3. Savant-Bhonsale, S. & Montell, D., J. *torso-like* encodes the localized determinant of Drosophila terminal pattern formation. *Genes & Development* **7,** 2548-2555 (1993).
4. Liu, F., Morrison, A. H. & Gregor, T. Dynamic interpretation of maternal inputs by the Drosophila segmentation gene network. *Proceedings of the National Academy of Sciences* **110,** 6724–6729 (2013).
5. Grimm, O. *et al.* Torso RTK controls Capicua degradation by changing its subcellular localization. ***Development*** **139,** 3962-3968 (2012).
6. Irish, V., Lehmann, R. & Akam, M. The Drosophila posterior-group gene nanos functions by repressing hunchback activity. *Nature* **338,** 646–648 (1989).
7. Manu *et al.* Canalization of gene expression in the Drosophila blastoderm by gap gene cross regulation. *PLoS Biol* **7,** e49 (2009).
8. Papatsenko, D. & Levine, M. The Drosophila gap gene network is composed of two parallel toggle switches. *PLoS ONE* **6,** e21145 (2011).
9. Segal, E., Raveh-Sadka, T., Schroeder, M., Unnerstall, U. & Gaul, U. Predicting expression patterns from regulatory sequence in Drosophila segmentation. *Nature* **451,** 535–540 (2008).
10. Rivera-Pomar, R., Lu, X., Perrlmon, N., Taubert, H. & Jäckle, H. Activation of posterior gap gene expression in the *Drosophila* blastoderm. *Nature* **376,** 253–256 (1995).
11. Li, X.-Y. *et al.* Transcription Factors Bind Thousands of Active and Inactive Regions in the Drosophila Blastoderm. *PLoS Biol* **6,** e27 (2008).
12. Niessing, D., Blanke, S., & Jäckle, H. Bicoid associates with the 5'-cap-bound complex of *caudal* mRNA and represses translation. *Genes & Development* **16,** 2576–2582 (2002).
13. Manu *et al.* Canalization of gene expression and domain shifts in the *Drosophila* blastoderm by dynamical attractors. *PLoS Comput Biol* **5**, e1000303 (2009).
14. Pignoni, F., Steingrímsson, E., Lengyel, J. A. *bicoid* and the terminal system activate *tailless* expression in the early *Drosophila* embryo. *Development* **115,** 239–251 (1992).
15. Dubuis, J. O., Samanta, R. & Gregor, T. Accurate measurements of dynamics and reproducibility in small genetic networks. *Molecular Systems Biology* **9,** 1–15 (2013).
16. Petkova, M. D., Tkačik, G. & Gregor, T. Optimal decoding of informaiton from a genetic network. 1–19 (2016).

17. Surkova, S. *et al.* Quantitative dynamics and increased variability of segmentation gene expression in the *Drosophila Krüppel* and *knirps* mutants. *Developmental Biology* **376,** 99–112 (2013).
18. Eldon, E. D. & Pirrotta, A. Interactions of the *Drosophila* gap gene *giant* with maternal and zygotic pattern-forming genes. *Development* **111,** 367–378 (1991).
19. Kraut, R. & Levine, M. Mutually repressive interactions between the gap genes *giant* and *Krüppel* define middle body regions of the *Drosophila* embryo. *Development* **111,** 611–621 (1991).
20. Wu, X., Vakani, R. & Small, S. Two distinct mechanisms for different positioning of gene expression borders involving the *Drosophila* gap protein giant. *Development* **125,** 3765–3774 (1998).
21. Hülskamp, M., Pfeifle, C. & Tautz, D. A morphogenetic gradient of hunchback protein organizes the expression of the gap genes *Krüppel* and *knirps* in the early *Drosophila* embryo. *Nature* **346,** 577–580 (1990).
22. Ma, W., Trusina, A., El-Samad, H., Lim, W. A. & Tang, C. Defining network topologies that can achieve biochemical adaptation. *Cell* **138,** 760–773 (2009).
23. Turing, A. M. The chemical basis of morphogenesis. *Philosophical Transactions of the Royal Society of London. Series B, Biological Sciences* **237,** 37-72 (1952).