Parametric identification of concentration-dependent gene regulation networks

Qie Hu

Electrical Engineering and Computer Sciences
University of California at Berkeley

Technical Report No. UCB/EECS-2015-34
http://www.eecs.berkeley.edu/Pubs/TechRpts/2015/EECS-2015-34.html

May 1, 2015

Parametric identification of concentration-dependent gene regulation networks

Qie Hu, Anil Aswani, Soile VE Keränen, David W Knowles,
Mark D Biggin, Claire J Tomlin

Abstract

In mathematical modeling of genetic regulation networks, the classical assumption is that the direction of regulation by a regulator (activation or repression) on a specific target gene is fixed. However experimental studies have suggested that the direction of gene regulation might be concentration-dependent. Earlier modeling work assumed only specific transcription factors behave in this concentration-dependent manner. Nevertheless, based on our *in vivo* DNA occupancy measurements of transcription factors and target gene mRNA expressions, we propose that concentration-dependent activation or repression might be a more general phenomenon. More specifically, we assume a regulator is activating at low concentrations and repressing at high concentrations, and we use the regulation of *even-skipped* stripe 2 (*eve*2) during Stage 5 of *Drosophila* embryogenesis as a case study. Our general approach is to describe the dynamic evolution of target mRNA concentration using a set of parametric Ordinary Differential Equations, which incorporate target gene mRNA transcription and degradation, and transcription factor expression and *in vivo* DNA occupancies. We show our model is better at predicting regulator importance and the effect of transcription factor mutations than two commonly used monotone models, where each regulator either activates or represses the target gene, but not both. In addition, our model generates spatial-temporal maps of factor activity, highlighting the times and spatial locations at which different transcription factors might regulate target gene expression levels. Finally, we use our model to design mutation experiments that help us to verify our assumption of concentration-dependent regulation by determining whether transcription factors that are currently thought to either only activate or repress *eve*2 in fact act in a concentration-dependent manner.

1 Introduction

Precise control of spatial-temporal gene expression patterns by transcription factors plays a central role in metazoan development. Quantitative, predictive mathematical models of this process are attractive because they provide precise hypotheses about the behavior of the system which can then be tested experimentally. These models take as input the protein expression patterns of transcription factors and seek to predict the output RNA patterns that these proteins drive via specific *cis*-regulatory regions (CRRs). Two of the central questions that these models address are which transcription factors function on each CRR and do these proteins activate or repress transcription and if so with what shaped mathematical function?

To select which transcription factors regulate a given CRR, models to date have used either qualitative biological information, generally derived from molecular genetic experiments, or the existence of strong

correlations between input transcription factor protein expression and the output CRR driven mRNA pattern [1, 2]. Many of these models have in addition determined the level at which each transcription factor binds to the CRR using *in vitro* data for the affinity of the proteins for specific DNA sequences, provided from a Position Weight Matrix (PWM) [3, 4]; whereas the other models have not employed DNA binding data [1, 2, 5, 6].

The above models, however, are unlikely to accurately reflect the range of transcription factors bound to CRRs in cells or their levels of DNA occupancy. *In vivo* crosslinking studies establish that CRRs are each bound in cells by more sequence specific DNA binding proteins than the 3-6 typically included in previous models [7]. Several tens of proteins can be bound to a single CRR over a range of DNA occupancies that result in regulatory responses varying from strong to weak [7]. In addition, measured levels of *in vivo* DNA occupancy are poorly predicted by models that use PWM data alone because they do not take into account competitive inhibition by nucleosomes or cooperative DNA binding between regulatory proteins [4, 8]. Therefore, to better represent the situation *in vivo*, here we have established models that rely strongly on *in vivo* crosslinking data for a large number of the major regulators within a network. Our goal is to determine how accurately output patterns can be predicted given *in vivo* DNA binding data and what mathematical functions best describe the likely regulatory behavior of transcription factors in controlling CRR output.

Most spatio-temporal models have assumed that the direction of regulation by a transcription factor (activation or repression) on a specific target CRR is fixed [9, 10, 11, 12, 13, 14, 15]. For example, linear, Hill, and Exponential models as well as sign-definite Boolean networks have all been used to represent such monotonic regulation. A number of experimental studies [16, 17, 18, 19, 20] and several modeling efforts [21, 22, 2, 5], however, have suggested that the direction of gene regulation might depend on the concentration of the transcription factor in the cell. For example, three animal transcription factors have been shown to activate some targets at low concentration but to repress them at higher concentrations [16, 17, 18, 19, 20]. In addition, here we show that of the relationship between measured transcription factor DNA occupancy *in vivo* and the rate of change in transcription for a specific CRR is best fit by a gaussian function that implies that most transcription factors activate the CRR when they are bound at low levels yet repress it when they occupy it at higher levels. Therefore, we have compared models that use two of the most common monotonic functions, Hill and Exponential, and a Gaussian function that allows concentration dependent effects to determine which more accurately reflects the known properties of the system.

To develop our models, we have examined control of the *even-skipped* stripe 2 (*eve*2)CRR by a set of 14 regulators in the *Drosophila* blastoderm embryo. The *eve*2 CRR is one of the most thoroughly characterized developmental CRRs and the blastoderm network has been subject to extensive genetic screens that have identified the majority of transcription factors that convey spatial patterning information to it [30, 25, 26, 23]. As part of the Berkeley Drosophila Transcription Network Project (BDTNP), we have previously generated *in vivo* DNA binding data for 17 of the major regulators of spatial patterning in the *Drosophila* blastoderm [23] and have also established a cellular resolution, three dimensional VirtualEmbryo that quantitates the expression patterns of many transcription factors and their targets. To allow our proposed models, here we have extended the VirtualEmbryo to include the LacZ reporter gene expression pattern driven by the *eve*2 CRR.

In preliminary models that did not employ DNA binding data, we previously showed that the regulation of *eve* gene expression was more accurately captured by Ordinary Differential Equations (ODEs) that fit to the rate of change in mRNA expression than by static models that examined correlation between transcription

factor and target gene expressions at a fixed time point [20]. Therefore, here we have again adopted (parametric) ODEs and have posed the parameter estimation problem as the optimization of a squared loss function between the experimentally measured wildtype mRNA patterns and the model's predictions. Our results show that models based on *in vivo* DNA binding data for many transcription factors predict the measured spatial-temporal patterns driven by the *eve*2 CRR quite well. A Gaussian model that allows concentration dependent effects agrees better with experimental genetic data on how transcription factors control the *eve*2 CRR than does monotonic models and also agrees more closely with the relative importance of each transcription factor suggested by rank order of measured levels of DNA occupancy of the transcription factors *in vivo*. This result suggests that concentration dependent effects may be more common than previous studies suggest. Finally, we use our model to design mutation experiments that test our prediction of concentration-dependent regulation by determining whether transcription factors act in a concentration-dependent manner or not.

1.1 Model of transcription control

The *eve*2 CRR is most strongly patterned along the Anterior/Posterior (A/P) axis of the *Drosophila* blastoderm embryo. Genetic screens have identified nine transcription factors that initiate developmental patterning along this axis of the embryo: the A/P early transcription factors Bicoid (Bcd), Caudal (Cad), Dichaete (D), Giant (Gt), Hunchback (Hb), Huckebein (Hkb), Knirps (Kni), Krüppel (Kr), and Tailless (Tll). All nine of these regulators have been found to bind at moderate to high levels to the *eve*2 CRR *in vivo* [23]; and although classic molecular genetic studies have only shown that four of these proteins directly regulate this CRR (BCD, GT, HB and KR) [24, 30, 25, 26], there is no evidence that the other five proteins do not. Therefore, we have included all nine of the A/P early regulators in our models to explore the potential role of each factor. Genetic screens have also identified additional transcription factors that control developmental patterning either along the A/P axis at later times or along the Dorsal/Ventral (D/V) axis. A number of these additional regulators also bind to the *eve*2 CRR *in vivo*, though generally at lower levels than the A/P early regulators [23]. D/V regulators have been shown to have weak ($\sim 30\%$) effects on the expression levels of *eve* stripe 2 [27] and a modest role for late A/P regulators cannot be ruled out. Therefore, we have included five of these further regulators in our models so that they more realistically represent the complexity of transcription factor DNA binding *in vivo* and capture weak regulatory events to which this CRR is potentially subject: Dorsal (Dl), Paired (Prd) and Runt (Run), Snail (SNA), and Twist (TWI).

To model the control of the *eve*2 CRR by these 14 transcription factors, we first measured expression for all 14 of these proteins and for the LacZ reporter mRNA driven by this CRR at four consecutive ∼ten minute intervals during the blastoderm stage. Expression was measured by multiphoton microscopy and the resulting image data was transferred into our VirtualEmbryo to quantitate expression in each of the $\sim 6,000$ cells of the embryo at each interval (e.g. Fig. 1). The transcription factor protein expression data was then used to calculate the level of each protein binding to the *eve*2 CRR in each cell by disaggregating chromatin immunoprecitation with microarray technology (ChIP-chip) DNA binding measurements obtained from whole embryos (Fig. 1). The ChIP-chip data for all transcription factors was first normalized such that the most highly bound regions in the genome had the same score, and thus the scores on the *eve*2 CRR reflect the relative strength of binding of each protein versus its most strongly controlled target genes. Our disaggregation of the ChIP-chip scores on the *eve*2 CRR assumes that for a given transcription factor, DNA occupancy in each cell is proportional to the protein's expression (Fig. 1). Previous modeling of *in vivo* DNA

binding suggests this assumption is a reasonable approximation as the chromatin accessibility is similar in most cells at this stage of development and interactions between transcription factors, which could modify occupancy differently in different subsets or cells, is limited [4].

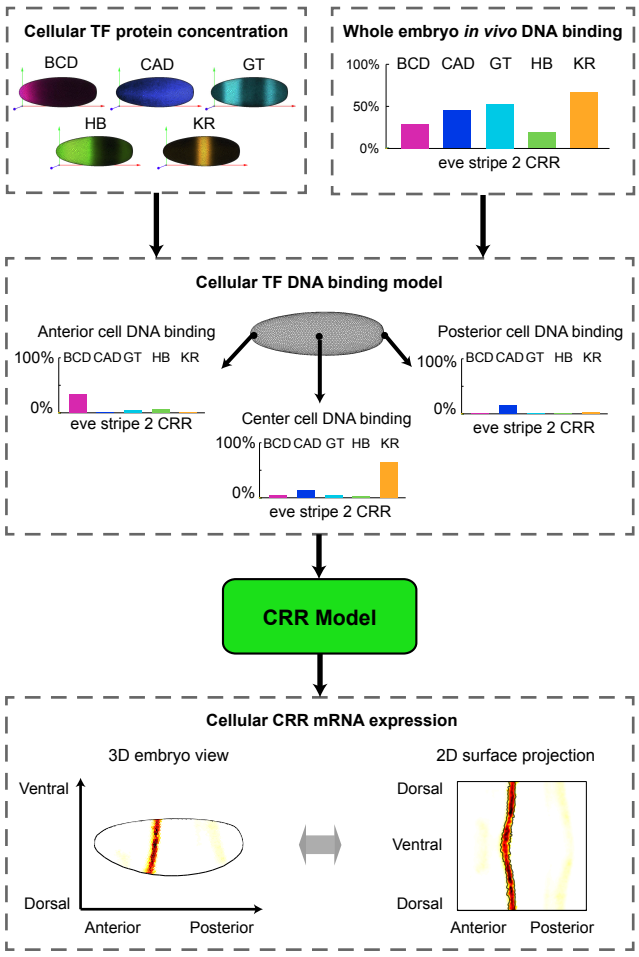

Figure 1: Overview of modeling approach. Flow diagram showing model input and output for a simplified version of the *eve*2 CRR with 5 transcription factor inputs. Cellular protein concentration of each transcription factor (top left) is used to disaggregate the whole embryo *in vivo* ChIP-chip measurements of DNA binding of that transcription factor (top right) on the *eve*2 CRR into each cell of the embryo. This computation is shown for one particular time point and three cells located anteriorly, centrally and posteriorly on the lateral side of the embryo (middle panel). Outcome of the model is a spatial-temporal expression pattern of *eve*2 mRNA (bottom panel). Result at one particular time point is shown here. We show gene expression and protein concentration data in two formats: as a three-dimensional embryo or as a two-dimensional rectangular projection of the unrolled surface of an embryo. All three-dimensional embryo images in this paper are shown with anterior to the left, posterior to the right, dorsal to the bottom and ventral to the top. All two-dimensional projections of embryos are shown with anterior to the left, posterior to the right, ventral in the middle and dorsal in the top and bottom.

Our model then takes these spatial-temporal patterns of DNA binding as input and predicts the spatial-temporal expression pattern that the CRR produces. The transcriptional process is modeled by a set of ODEs that takes the expression pattern of *eve*2 mRNA at the first time point Stage 5:9-25 as an initial condition to simulate *eve*2 mRNA pattern for all subsequent time points.

1.2 Gaussian relationship between transcriptional rate and DNA binding

Since LacZ mRNA has a half-life of less than 6 min, which is shorter than the time scale of gene expression changes, it is plausible to assume that the expression of mRNA is proportional to the transcriptional rate [3]. We also approximated the DNA binding of transcription factors as constant between successive measurement time points. Then, to determine the relationship between transcriptional rate and the level of DNA binding of transcription factors, we plotted the change in LacZ mRNA expression in each cell between two successive time points versus the binding of each factor in that cell during this time interval (Fig. 2). We found that while some transcription factors, such as Hb, showed a monotone correlation between the level of transcription factor binding and the change in transcriptional output between subsequent time points, many other transcription factors showed a more complex relationship. It appeared that many transcription factors activate the *eve*2 CRM when they are bound at low binding levels and repress this same CRM at higher binding levels.

Thus, we performed hypothesis testing for monotonicity on each transcription factor's regulation data, using the methods described in [28]. The null hypotheses are that the relationship between transcriptional rate and the level of DNA binding of a transcription factor is either monotonically increasing or monotonically decreasing, which corresponds to transcription factors being either activators or repressors. Corrections for multiple testing were made using weighted Bonferroni, where the weights are the transcription factor ChIP-chip scores. Corrected p-values for each hypothesis test are shown in Table S1. We found that at a 5% significance level, the null hypotheses are rejected for Kr and Gt. In other words, for these two transcription factors, the correlation data support the idea that they act in a concentration dependent manner. We do not feel at this stage one can rule out the possibility that some of other transcription factor may also act in a concentration dependent manner given the apparent 'up' and 'down' relationships seen in Fig. 2.

As a result, for each regulator, we plotted three functions that each minimized the sum of squared error between their functional values and the experimentally measured changes in mRNA expression. The monotone functions, Exponential and Hill, represent models where transcription factors are either activators or repressors, whereas the Gaussian function can be used to describe concentration-dependent activation or repression by transcription factors. Performances of the three functions were compared using weighted-root-mean-square-error (E_{w-RMS}). For each function, we computed its E_{w-RMS} by first computing the square root of the sum of squared fitting errors for each transcription factor and then, a weighted average across all 14 transcription factors using factors' ChIP-chip scores as weights. We found that the Gaussian function's E_{w-RMS} was marginally less than those of the two monotone models (Table 1). This suggested that Gaussian functions might be a reasonable candidate for describing the rate of transcriptional regulation when controlled by a set of factors bound to the CRR.

2 Mathematical Model

2.1 CRR binding data

We compute the CRR-binding levels of transcription factors in each cell using cellular resolution transcription factor proteins' expression data and embryo resolution ChIP-chip binding data. Let $x_{i,j}(t) \in \mathbb{R}_+$ and $c_{i,j}(t) \in \mathbb{R}_+$ represent the CRR-binding level and protein expression of the i^{th} transcription factor in the j^{th} cell at

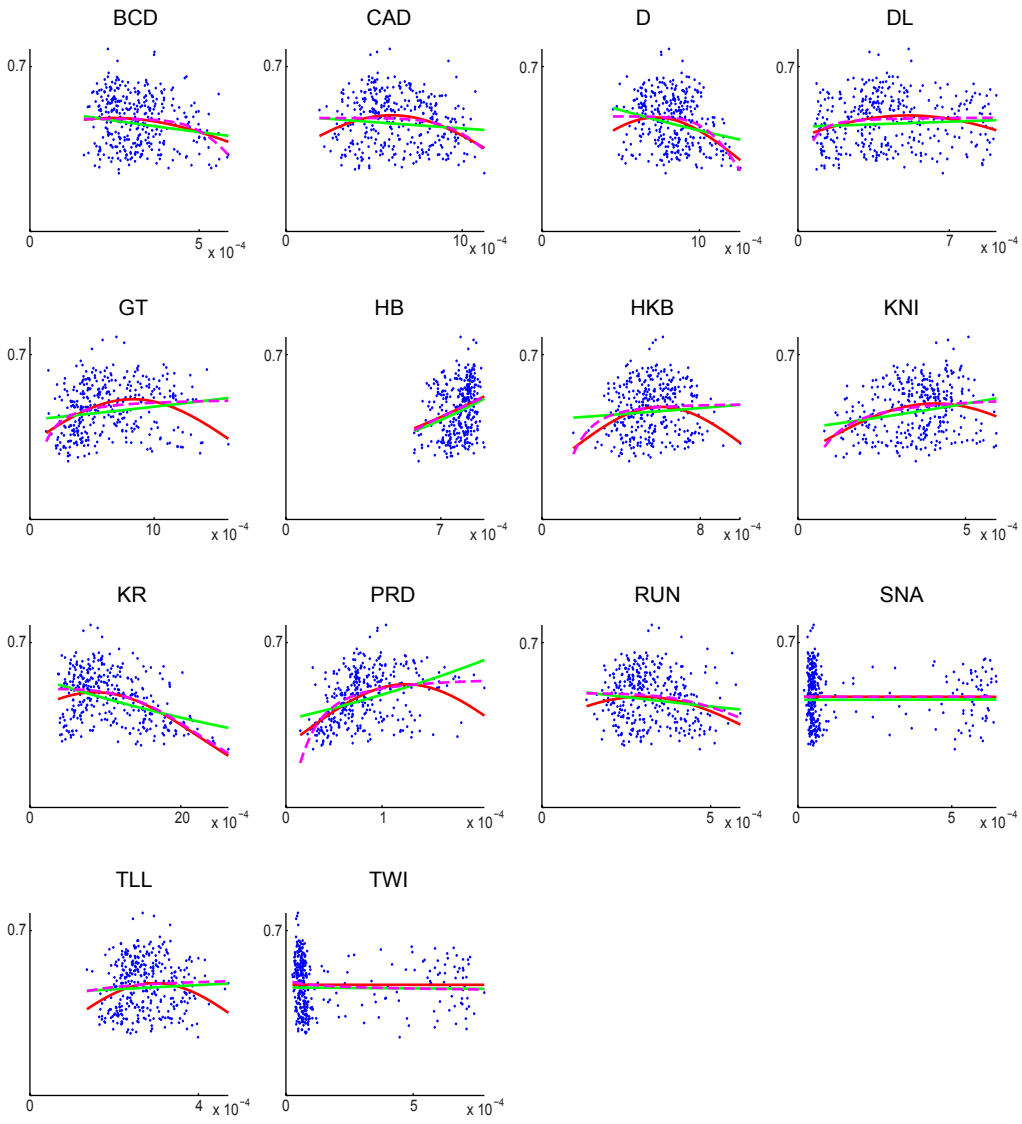

Figure 2: Change in mRNA expression versus DNA binding of transcription factors. Horizontal axis represents the DNA binding level of a transcription factor in one cell at time t. Vertical axis represents the change in mRNA expression in the same cell between time t and $t+1$. Each blue data point corresponds to one cell in the embryo between two given measurement time points, and are taken from experimental data. Solid red curve represents the Gaussian function that minimizes the RMS error, in the vertical direction, between the curve and experimental data points. Solid green curve represents the Exponential function that minimizes the same RMS error. Dashed magenta curve represents the Hill function that minimizes the same RMS error.

Table 1: Comparison of best fit Gaussian, Exponential and Hill functions to change in LacZ mRNA expression versus transcription factor binding data in Fig. **??**

	Gaussian	Exponential	Hill
E_{w-RMS}	0.2627	0.2667	0.2639
% increase in E_{w-RMS} compared to Gaussian	-	1.6	0.5

Table shows the root-mean-square (RMS) error for each function; weighted by transcription factor ChIP-chip scores (E_{w-RMS}), and the percentage increase in E_{w-RMS} of the Exponential and Hill functions compared to the Gaussian function, respectively. $E_{w-RMS} = \frac{\sum_i \left(w_i || \bar{\delta y_i} - \delta y_i ||\right)}{\sum_i w_i}$, where $\bar{\delta y_i}$ is the vector of measured change in LacZ mRNA expression between 2 successive time points in each cell for the i^{th} transcription factor, δy_i is the same change in LacZ mRNA expression predicted by Gaussian/Exponential/Hill functions and w_i is the ChIP-chip score of the i^{th} transcription factor.

time t, respectively. In addition, we use b_i to represent ChIP-chip score of the i^{th} regulator. Finally we denote the time that ChIP-chip measurements are taken as τ. The binding level of the i^{th} transcription factor in the j^{th} cell at time t is then approximated using

$$x_{i,j}(t) = \frac{b_i c_{i,j}(t)}{\sum_{j=1}^{J} c_{i,j}(\tau)} \tag{1}$$

for $i = 1, \ldots, I$ and $j = 1, \ldots, J$, where I is the number of transcription regulators and J is the number of cells in the embryo. This method follows the law of mass action applied to the elementary reaction of binding between regulator proteins and nucleotides on the CRR. This binding model has been successfully applied by others [4, 49].

2.2 CRR transcription model

We model the regulation of *eve*2 expression by 14 transcription factors: Bicoid, Caudal, Dichaete, Dorsal, Giant, Hunchback, Huckebein, Knirps, Krppel, Paired, Runt, Snail, Tailless and Twist. We define a vector $y(t) := [y_1(t), y_2(t), \ldots, y_J(t)]^T \in \mathbb{R}_+^J$, whose components represent the concentration of *eve*2 mRNA in a single cell, and use a parametric model where we have a group of Ordinary Differential Equations (ODEs) with unidentified parameters. Each ODE describes the dynamics of *eve*2 mRNA concentration y_j inside a single cell:

$$\frac{dy_j}{dt} = k \prod_{i=1}^{I} f_i(x_{i,j}) - \lambda y_j, \tag{2}$$

where $x_{i,j}$ is as defined in Sec. 2.1 and $k, \lambda \in \mathbb{R}_+$ are unknown parameters. Furthermore, f_i represents the regulation function for the i^{th} transcription factor. Its mathematical form depends on the transcription model under consideration:

$$
\begin{aligned}
\text{Gaussian model:} \quad & f_i(x_{i,j}) = \exp\left(-\frac{(x_{i,j}-\rho_i)^2}{\omega}\right), \\
\text{Exponential model:} \quad & f_i(x_{i,j}) = \exp\left(\rho_i x_{i,j}\right), \\
\text{Hill model:} \quad & f_i(x_{i,j}) = 1_A(i)\left(\frac{x_{i,j}^{\omega}}{\rho_i^{\omega}+x_{i,j}^{\omega}}\right) + 1_R(i)\left(1-\frac{x_{i,j}^{\omega}}{\rho_i^{\omega}+x_{i,j}^{\omega}}\right),
\end{aligned} \tag{3}
$$

where $\rho_i \in \mathbb{R}$, $i = 1, 2, \ldots, I$ and $\omega \in \mathbb{R}_+$ are further unknown parameters, A is the set of activators and R is the set of repressors. In addition, $1_\chi(\cdot) : \{1, 2, \ldots, I\} \to \{0, 1\}$ is the indicator function where $1_\chi(i) = 1$ if $i \in \chi$ and $1_\chi(i) = 0$ if $i \notin \chi$. The first term in Eq.2 represents the effect of transcription factor regulation on *eve*2 and the second term represents mRNA degradation.

The sign of the partial derivative of the regulation function f_i with respect to the binding level $x_{i,j}$ and evaluated at a specific binding $x_{i,j}^*$, $\frac{\partial f_i}{\partial x_{i,j}}(x_{i,j}^*)$, indicates the regulatory direction of transcription factor i in cell j at the binding level $x_{i,j}^*$. If $\frac{\partial f_i}{\partial x_{i,j}}(x_{i,j}^*) > 0$, then for a unit increase in the binding level of transcription factor i, keeping all remaining transcription factors' binding unchanged, the value of the function f_i will increase. Hence the rate of transcription of the mRNA, $\frac{dy_j}{dt}$, will also increase. This corresponds to an activating effect. On the other hand if $\frac{\partial f_i}{\partial x_{i,j}}(x_{i,j}^*) < 0$, then for a unit increase in the binding level of the i^{th} transcription factor, f_i and hence $\frac{dy_j}{dt}$ will decrease. This corresponds to repressive effect. The magnitude of this partial derivative represents the strength of regulation. Note that the direction and strength of regulation depends both on the function f_i and the binding level $x_{i,j}^*$.

For monotonic models, the regulation function f_i is such that, given i and j, $\frac{\partial f_i}{\partial x_{i,j}}(x_{i,j}^*)$ is of the same sign for all $x_{i,j}^* \geq 0$, $i = 1, 2, \ldots, I$ and $j = 1, 2, \ldots, J$. For the Exponential model,

$$
\frac{\partial f_i}{\partial x_{i,j}}(x_{i,j}^*) = \rho_i \exp\left(\rho_i x_{i,j}^*\right). \tag{4}
$$

Therefore if transcription factor i's corresponding parameter ρ_i is positive, then Eq. 4 is positive for all $x_{i,j}^* \geq 0$. In other words, transcription factor i always activates the transcription of the target gene. On the other hand, if ρ_i is negative for transcription factor i, then Eq. 4 is negative for all $x_{i,j}^* \geq 0$, and hence the transcription factor is always a repressor. Similarly for the Hill model. For an activator,

$$
f_i(x_{i,j}) = \frac{x_{i,j}^{\omega}}{\rho_i^{\omega}+x_{i,j}^{\omega}} \tag{5}
$$

and its partial derivative is

$$
\frac{\partial f_i}{\partial x_{i,j}}(x_{i,j}^*) = \frac{\omega \rho_i^{\omega} x_{i,j}^{*\omega-1}}{\left(\rho_i^{\omega}+x_{i,j}^{*omega}\right)^2} \geq 0, \quad \forall x_{i,j}^* \geq 0. \tag{6}
$$

For a repressor, the regulation function is

$$
f_i(x_{i,j}) = \frac{\rho_i^{\omega}}{\rho_i^{\omega}+x_{i,j}^{\omega}} \tag{7}
$$

and its partial derivative is

$$\frac{\partial f_i}{\partial x_{i,j}}(x^*_{i,j}) = \frac{-\omega\rho_i^\omega x^{*\omega-1}_{i,j}}{\left(\rho_i^\omega + x^{*\omega}_{i,j}\right)^2} \leq 0, \quad \forall x^*_{i,j} \geq 0. \tag{8}$$

Therefore in monotonic models, transcription factors either always activate or always repress the transcription of the target gene, despite its binding level to the DNA.

In the Gaussian model, whether a transcription factor is activating or repressing the target gene depends on its binding level. Based on our data, we assume it to be generally activating at low binding levels and repressing at high binding levels. More specifically, $\frac{\partial f_i}{\partial x_{i,j}}(x^*_{i,j})$ is positive when $x^*_{i,j} < \rho_i$, corresponding to up-regulation; and $\frac{\partial f_i}{\partial x_{i,j}}(x^*_{i,j})$ is negative when $x^*_{i,j} > \rho_i$, corresponding to down-regulation.

2.3 Nonlinear least squares formulation

We define a vector θ, whose components are the parameters of our model. For both the Gaussian and the Hill models $\theta := [k, \lambda, \omega, \rho_1, \ldots, \rho_I]^T \in \{\mathbb{R}^3_+ \times \mathbb{R}^I\}$, and for the Exponential model $\theta := [k, \lambda, \rho_1, \ldots, \rho_I]^T \in \{\mathbb{R}^2_+ \times \mathbb{R}^I\}$.

The parameters are estimated by choosing the set of parameter values with which the model best describes the experimental data. In other words, we would like to minimize the error between our model's output and the data. Recall that our experimental data is a sequence of *eve*2 mRNA concentrations measured at four time points, $t_0, t_1, \ldots, t_3$, throughout stage 5 of embryogenesis. Mathematically, this is posed as an optimization problem

$$\begin{aligned} \arg\min_\theta \; & J(\theta) = \textstyle\sum_{n=1}^3 \|y_\theta(t_n) - \tilde{y}(t_n)\|^2 \\ \text{subject to } & y_\theta(t_0) = \tilde{y}(t_0) \\ & \tfrac{dy_{\theta,j}}{dt} = k\prod_{i=1}^I f_{\theta,i}(x_{i,j}) - \lambda y_{\theta,j}, \; \forall j \\ & k, \lambda, \omega > 0 \end{aligned} \tag{9}$$

where $y_\theta(t)$ is the vector solution to the set of ODE constraints in Eq. 9 with parameters θ and $\tilde{y}(t)$ is the set of experimentally measured wildtype *eve*2 mRNA concentration values from the VirtualEmbryo that has been processed as described in Sec. 1.1.

2.4 Numerical implementation and algorithms

Values of the function $f_{\theta,i}$ can be very small for certain combinations of binding concentrations and parameter values. Repeated multiplication of small numbers may then cause arithmetic underflow of the product or arithmetic overflow of the scaling parameter k. To overcome this, we define $\hat{k} := \ln(k)$ and rewrite the ODE constraints in Eq. 9 as

$$\frac{dy_{\theta,j}}{dt} = \exp\left(\hat{k} - \frac{1}{\omega}\textstyle\sum_{i=1}^I (x_{i,j} - \rho_i)^2\right) - \lambda y_{\theta,j}. \tag{10}$$

We solved this constrained nonlinear (and nonconvex) optimization problem using the MATLAB function *fmincon* with *trust-region-reflective* algorithm. The approximated Hessian matrix is computed using finite differences.

We computed the solution of the ODEs, y_θ, and the gradient of the objective function, $\nabla_\theta J$, analytically by approximating the binding concentrations of transcription factors as constant between successive measurement time points. With this approximation, the first term in Eq. 10 became a constant for all $t \in [t_n, t_{n+1})$, which we denoted $\alpha_{\theta,j,n}$. Therefore the system became piecewise linear and time invariant:

$$\frac{dy_{\theta,j}}{dt} = \alpha_{\theta,j,n} - \lambda y_{\theta,j}, \tag{11}$$

for $t \in [t_n, t_{n+1})$ and $n = 0, 1, 2, 3$.

This allows us to easily write down an analytical solution for $y_\theta(t_n)$ as a function of $y_\theta(t_0)$, which is known:

$$y_\theta(t_n) = \left[\prod_{l=0}^{n-1} A_{\theta,n}\right] y_\theta(t_0) + \sum_{m=0}^{n-2} \left[\prod_{l=m+1}^{n-1} A_{\theta,l}\right] B_{\theta,m} + B_{\theta,n-1}, \tag{12}$$

where

$$\begin{aligned} A_{\theta,n} &= \exp(-\lambda(t_n - t_{n-1})) \in (0,1), \\ B_{\theta,n} &= \lambda^{-1}(1 - A_{\theta,n})\,\alpha_{\theta,n} \in \mathbb{R}_+^J, \\ \alpha_{\theta,n} &= [\alpha_{\theta,1,n}, \ldots, \alpha_{\theta,J,n}]^T \in \mathbb{R}^J. \end{aligned} \tag{13}$$

In addition the gradient of the objective function is given by

$$\nabla_\theta J = 2 \sum_{n=1}^{3} \nabla_\theta y_\theta(t_n)^T \left(y_\theta(t_n) - \tilde{y}(t_n)\right). \tag{14}$$

Since the problem is nonconvex, and gradient-based search algorithms only guarantee convergence to local minima, the *fmincon* function is repeatedly run for different initial parameter guesses in order to find the global minimum.

3 Results

3.1 Model fit to training data: wildtype *eve*2 mRNA

We assessed the fit of our model to the training data, wildtype *eve*2 mRNA pattern, both qualitatively and quantitatively. Experimentally measured *eve*2 stripe, regulated by the early *eve*2 CRM in this study, weakens towards the end of Stage 5, and the stripe is stronger in the lateral and dorsal regions than it is in the ventral region (Fig. 3). Simulation results from our Gaussian model match these qualitative experimental observations.

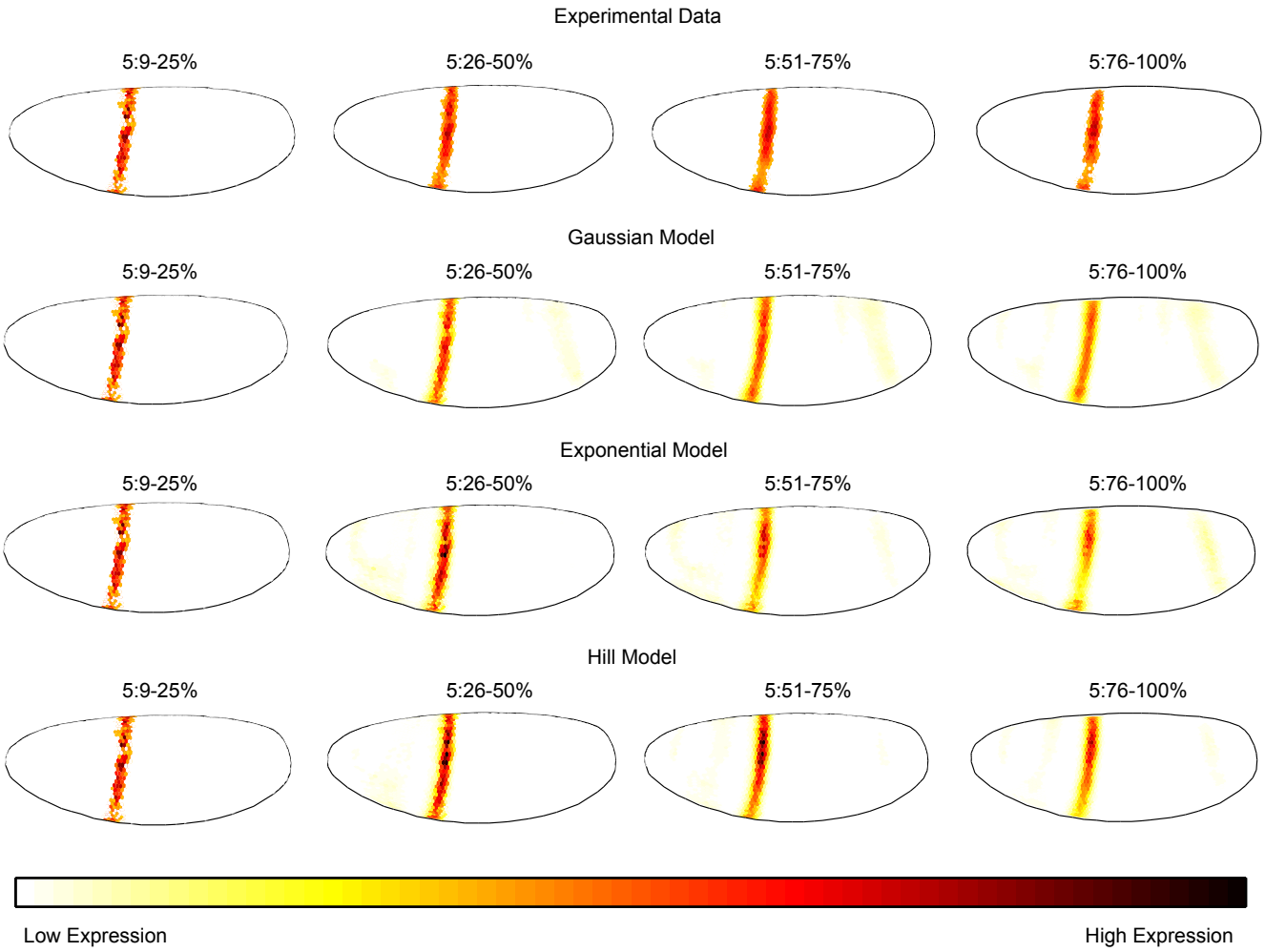

Figure 3: Comparison of model predictions for the training data. The top row of embryos represents the training data - experimentally measured wildtype *eve*2 mRNA patterns at 4 time points 5:9-25%, 5:26-50%, 5:51-75% and 5:76-100%. The bottom three rows represent the predictions by the Gaussian, Exponential and Hill models respectively. The color of each cell in the embryo represents mRNA expression. Red/black colored cells correspond to higher mRNA expression. Yellow/white colored cells correspond to low/no mRNA expression. The mRNA expressions in all embryos are plotted on the same scale. Note that the *eve*2 pattern at initial time point 5:9-25% is identical for all models and the experimental data. This is because all dynamic models use the experimental *eve*2 pattern at time point 5:9-25% as the initial condition for the ODEs.

Table 2: Mean-Squared-Error (MSE) of each model's fit to training data (wildtype *eve*2 pattern)

Gaussian	Exponential	Hill
26.74	34.64	25.55

Quantitatively, we measured the goodness-of-fit of the model to the training data using mean-squared-error. Mean-squared-error (MSE) is a commonly used statistical measure of how well model fits experimental data, where a lower value of MSE qualitatively means that a model better fits the data. Table 2 shows the MSE of the Gaussian model. Our solution shows low levels of *eve*2 expression in the anterior and posterior of the embryo. This is probably because our model only includes 14 out of more than 21 transcription factors bound to the *eve*2 CRR at this stage of embryogenesis [23].

In addition, our model is able to produce *eve*2 mRNA pattern *de Novo*. In other words, we first trained the model using a modified training set that contains *eve*2 mRNA pattern at an additional earlier time point Stage 5:5-8, which is zero throughout the whole embryo. Then we simulate the learnt model using the set of normal transcription factor expressions as inputs and zero mRNA expression as initial condition for the ODEs (Fig. S1).

3.2 Effect of ChIP-chip measurement noise

To validate our model, we investigated how noise in ChIP-chip binding measurements affects our model's fit to training data. More specifically, during the training stage, we added zero mean Gaussian distributed random measurement noise to the ChIP-chip binding data, while keeping both transcription factor and LacZ mRNA expression data unchanged. We then simulated the trained model using original binding data (without noise) and transcription factor expression data, and computed the MSE between experimentally measured mRNA patterns and the model's results. For each standard deviation of the measurement noise, this training and testing process was repeated 10 times and the average MSE was recorded.

When the magnitude of measurement noise is small (standard deviation up to ~ 0.1), the mean squared prediction error does not increase significantly (less than approximately 10%) (Fig. 4). As noise increases further, the model's prediction error increases rapidly. This result provides the first piece of evidence that our model is not over-fit, as it does not have enough parameters to fit it to noise.

3.3 Model prediction of relative strength of transcription factors

We were also able to determine the relative strengths of transcription factors in the formation of *eve*2 stripe pattern predicted by the model. To do this, we simulated mutant embryos *in silico*. Take Gt for example, we changed the ChIP-chip binding for Gt to the CRR to α% of its measured (wildtype) value, and kept the binding of all other transcription factors unchanged. We chose $\alpha = 30$ for all 14 transcription factors. We then used the same disaggregation method to re-compute Gt's binding level in each cell of the mutant embryo and simulated the mRNA pattern for this Gt mutant. The MSE between this simulation result and experimentally measured wildtype mRNA patterns was computed. This MSE is larger than that between our model's simulation results for the wildtype embryo and the experimentally measured wildtype mRNA patterns. In particular, a larger increase in MSE from the wildtype simulation results to the mutant results indicates a transcription factor with stronger regulation power, since its binding levels to the CRR has a

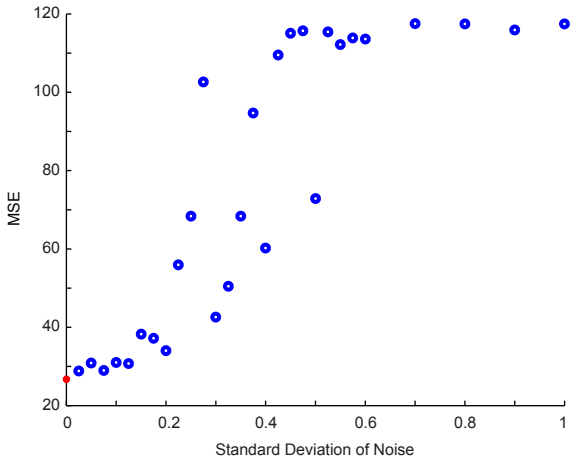

Figure 4: Mean-squared-error (MSE) of Gaussian model versus standard deviation of ChIP-chip measurement noise. For each noise standard deviation, independent zero-mean Gaussian noise is added to every transcription factor's ChIP-chip binding measurement. The Gaussian model is trained using this noise corrupted ChIP-chip data. The trained model is then used to reproduce the training data - wildtype *eve*2 pattern, using un-corrupted ChIP-chip binding and transcription factor expression data. The MSE between the model's prediction and the experimentally measured pattern is computed. For each noise standard deviation, this process is repeated ten times and the average value of MSE is plotted here.

larger effect on the resulting transcriptional output. We then ranked all 14 regulators in order of decreasing importance. This rank list closely matched the rank order of ChIP-chip scores of the transcription factors (Table 3).

3.4 Model predictions of mutation experiments

Our model was trained using wildtype *eve*2 mRNA patterns. For model validation, we used our model to predict genetic mutant patterns of *eve*2 mRNA. Published experimental results show that decreasing the binding of Gt causes *eve* stripe 2 to broaden anteriorly, whereas decreasing the binding of Kr causes the stripe to broaden into the posterior of the embryo [30, 25, 26, 40, 41]. When Bcd's or Hb's binding to *eve*2 CRR is reduced, weaker stripe intensities are experimentally observed [29, 30, 25, 26, 41]. Our model's predictions correctly capture all the distinguishing features of these four mutations (Fig. 5 shows the mutant results for Kr and Gt, Fig. S3, S4, S5 and S6 show results for all four mutants).

3.5 Factor activity plots

The model generated by our technique can be visualized as spatial-temporal maps of factor activities. These maps show the strength and variation of predicted regulation effects of transcription factors at different positions in the embryo and at different time points. More specifically they plot the amount of change in the rate of mRNA transcription caused by a unit increase in the transcription factor's binding. An example is shown in Fig. 6 for Stage 5:9-25. It shows the regulation effect of 4 transcription factors (directly or indirectly) on the formation of *eve* stripe 2. Blue values indicate marginal repressive effects (i.e. a unit increase in the regulators binding will cause a reduction in the rate of *eve* mRNA formation), and yellow/red

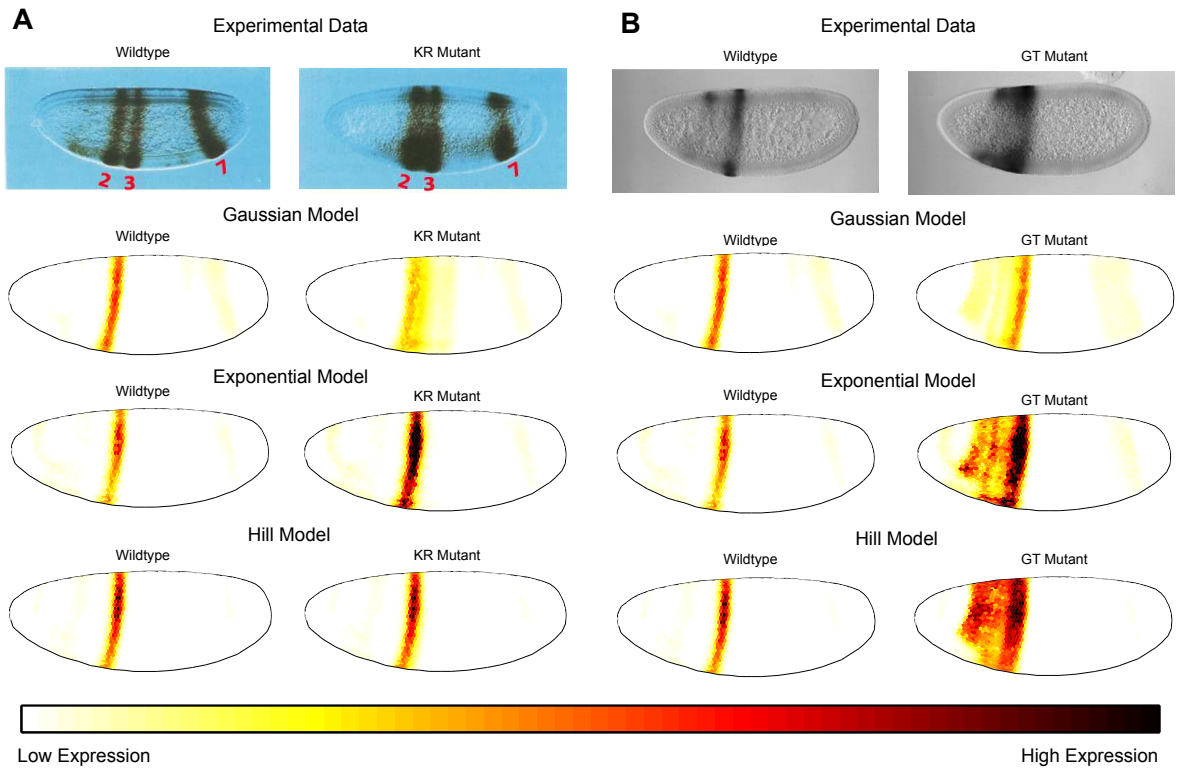

Figure 5: Comparison of model predictions for Gt and Kr mutants. For both (A) and (B), the first row shows experimental images of the wildtype embryo (left) and mutant embryo (right) from the literature, the bottom three rows show predictions of the wildtype embryo (left) and mutant embryo (right) by the Gaussian, Exponential and Hill models respectively. (A) Kr mutation. Experimental images are from [40], and show *eve* stripes 2, 3 and 7. Experimentally, stripe 2 expands posteriorly and fuses with stripe 3 in the Kr mutant embryo. The Gaussian model correctly predicts this posterior expansion of stripe 2 in the mutant. Both the Exponential and Hill models are not able to produce this. (B) Gt mutation. Experimental images are from [30]. Only *eve* stripe 2 is shown. Experimentally, stripe 2 expands anteriorly in the Gt mutant embryo. All three models correctly predict this mutant behavior.

Table 3: Comparison of transcription factor ranking as given by ChIP-chip scores and that predicted by each model

Rank	ChIP-chip	Gaussian	Exponential	Hill
1	Kr	Kr	Gt	Gt
2	Gt	Gt	Twi	Hb
3	Kni	Hb	Kni	Sna
4	Hkb	D	Kr	Kni
5	Cad	Kni	Hb	D
6	D	Cad	Dl	Prd
7	Tll	Hkb	Bcd	Kr
8	Bcd	Dl	D	Bcd
9	Run	Bcd	Tll	Tll
10	Dl	Tll	Sna	Cad
11	Hb	Run	Hkb	Run
12	Twi	Twi	Cad	Dl
13	Sna	Sna	Prd	Hkb
14	Prd	Prd	Run	Twi

Transcription factors ranked higher in the list correspond to stronger regulators of *eve*2 gene.

values correspond to marginal activating effects (i.e. a unit increase in the regulators binding will cause an increase in the rate of *eve* mRNA formation).

3.6 Comparison to monotone models

To help the understanding of our Gaussian model and establish its usefulness, we compared it with two commonly used monotone models: Exponential [3] and Hill [9, 11] model. The performance of three models is compared under three metrics: MSE, regulator importance ranking and mutation predictions. The MSE measures the goodness-of-fit of each model to the training data, and a lower value of MSE qualitatively means that a model better fits the training data. We used Spearmans Rank Correlation Coefficient (SRCC) to measure the linear correlation between the ranked list of transcription factor strengths predicted by a model and the rank order of ChIP-chip scores of the transcription factors. SRCC values closer to 1 are better. Finally for mutation predictions, we compared all three models predictions for Kr, Gt, Bcd and Hb mutants with published experimental results.

The Exponential model's MSE was 30% larger than that of the Gaussian model, indicating the former model has a poorer fit to the training data (Table 2). On the other hand, the Hill model appears to better fit the training data than the Gaussian model with a 4% less MSE (Table 2). However, examining the models' predicted rank order of regulators (Table 4), it is clear that those predicted by the monotone models correlate less well with the rank order given by the ChIP-chip scores, than the order predicted by the Gaussian model. Quantitatively, the Gaussian model had a SRCC of 0.78 with the ChIP-chip rank order, whereas the Exponential and Hill model's SRCC values were 0.33 and 0.10, respectively. Finally, all three models correctly predicted weakening in *eve* stripe 2's intensity for Bcd and Hb mutants (Fig. S3, S4), and an anteriorly broadened stripe for Gt mutants (Fig. 5). However both monotone models were not able to predict the posterior expansion of *eve* stripe 2 for a Kr mutant, whereas the Gaussian model was able to (Fig. 5).

To summarize, the Gaussian model outperformed the Exponential model under all three metrics consid-

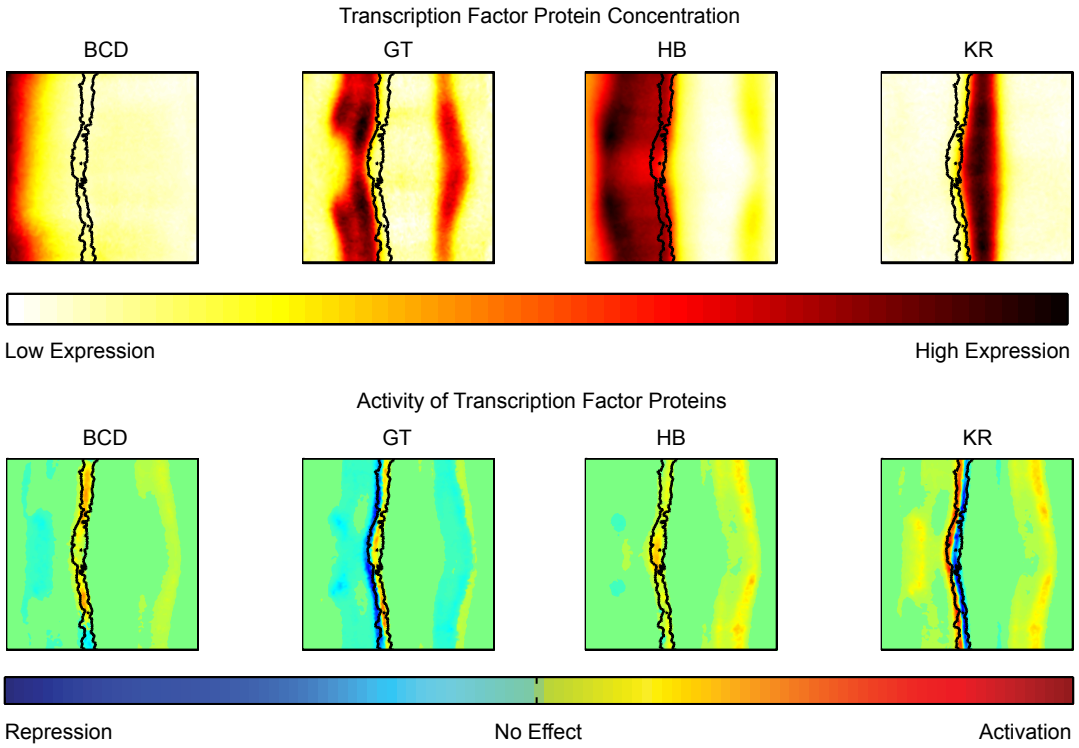

Figure 6: Transcription factor activity plots. Protein concentrations of four regulators (Bcd, Gt, Hb, Kr) and their activities (activation or repression) at time point 5:9-25%. Solid black lines delineate the boundary of *eve* stripe 2. The first row represents protein concentration. Color of cells represents concentration of proteins. Red/black colored cells correspond to higher concentrations. Yellow/white colored cells correspond to little/no protein concentration. The second row represents the regulation activity of each factor. Color of cells represents the direction and magnitude of regulatory activity. Blue cells indicate repressive effects. Yellow/red cells correspond to activating effects.

Table 4: Comparison of Spearman's Rank Correlation Coefficient (SRCC) between the rank order of ChIP-chip scores of the transcription factors and the rank order predicted by each model

Gaussian	Exponential	Hill
0.78	0.33	0.10

SRCC = 1 corresponds to perfect match between the rank order of ChIP-chip scores and the rank order predicted by a model, thus SRCC values closer to 1 are better.

ered; whereas the Hill model appeared to have poor predictive power despite being the best fit, amongst the three models, to the training data. Thus the Hill model seemed to be over-fit to the training data. Consequently, our systems identification results lend evidence that the Gaussian model may more accurately describe gene regulation than the commonly used monotonic Exponential and Hill models.

We repeated above comparison by adding a general ubiquitous activator to the set of transcription regulators [24], and we found that this did not affect the comparison results between the Gaussian and the monotone models. Details of this comparison are contained in Appendices A and B.

4 Discussion

Using our concentration-dependent Gaussian model, we made some genetic mutation predictions (Fig. 7) that can be implemented experimentally to test our hypothesis of concentration-dependent regulation, by determining whether transcription factors that are currently thought to either only activate or repress *eve*2 in fact act in a concentration-dependent manner.

For example, Kr is commonly described as a repressor and functions to establish the posterior boundary of *eve* stripe 2 [30, 25, 26, 40, 41]. Our model also predicted that Kr is a repressor for *eve*2 in the posterior region of the stripe. Yet it is also predicted to be an activator for *eve*2 in the anterior region of the stripe where its binding levels are lower (Fig. 6). Similar results were predicted in [20], where a nonparametric modeling methodology was used. Thus according to our model, increasing the binding of Kr by 100% would decrease *eve* mRNA expression near the posterior of the stripe, but would increase its expression in the anterior region. This would result in a thinning and anterior shift of *eve*2 in the mutant.

The function of Bcd is not straightforward, despite the apparent agreement. The need for Bcd to successfully establish *eve*2 suggests it is an activator [29, 30, 25, 26, 40, 41, 42]. However, at the anterior tip of the embryo, where Bcd is most highly expressed, no expression of *eve*2 is observed and the two main repressors of *eve*2, Gt and Kr, are also present at low concentrations. Various studies have searched for *eve*2 repressors in this region and candidates such as Torso, Capicua, Sap18 and Slp1 have been identified [43, 44, 45, 46, 47], yet none of these studies have been conclusive. This apparent paradox could be reconciled if Bcd's regulation was concentration-dependent: it is an activator around stripe 2 where its binding is low, and is a repressor in the anterior of the embryo where its binding is high (Fig. 6). Numerous published results have suggested that Bcd may act as a repressor either directly or indirectly: Bcd contains a self-inhibitory domain that has been shown to repress its own activity in Drosophila tissue culture cells [48, 49]; proteins such as Torso, Capicua and Sap18 have been suggested to interact with Bcd to repress Bcd target genes [44, 45, 46, 47]. Furthermore, [2, 35] have shown in their modeling work that allowing concentration-dependent activation or repression for Bcd improves predictions for a number of Drosophila segmentation genes mRNA patterns. To experimentally verify the repressor activity from Bcd, our model predicts that lowering Bcd binding to 10% of its wildtype value would lead to an increase in expression of *eve* mRNA in the anterior of the embryo.

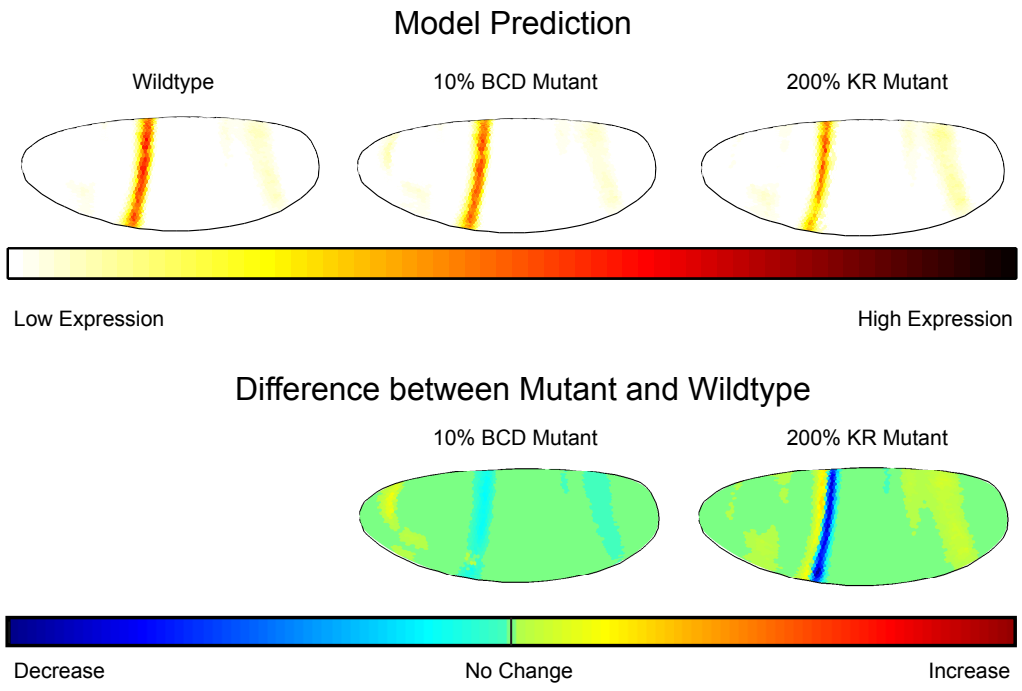

Figure 7: Concentration-dependent predictions. The first row shows the Gaussian model's prediction for *eve*2 mRNA pattern in the following three cases at Stage 5:51-75% : wildtype, mutant embryo with Bcd binding reduced to 10% of its wildtype value, mutant embryo with Kr binding increased to 200% of its wildtype value. Red/black colored cells correspond to higher concentrations. Yellow/white colored cells correspond to little/no protein concentration. The second row shows the difference in the *eve*2 mRNA expression in the mutant compared to the wildtype. Yellow/red cells indicate *eve*2 mRNA expression is higher in the mutant than in the wildtype. Blue cells indicate a decrease in mRNA expression in the mutant embryo.

A Supplementary Tables

Table S1: Weighted Bonferroni corrected P-values for monotone hypothesis test

Transcription factor	ChIP-chip score	P-value for H_1	P-value for H_2
Kr	6.4	0.001	0.003
Gt	5.2	0.001	0.001
Hkb	5.0	0.223	0.103
Kni	5.0	0.412	0.001
Cad	4.7	0.007	0.065
D	4.5	0.001	0.899
Tll	4.3	0.998	0.010
Bcd	3.2	0.448	1.000
Run	2.7	0.001	1.000
Dl	2.6	1.000	0.004
Hb	2.1	0.999	0.001
Twi	1.4	0.001	0.640
Sna	1.3	1.000	1.000
Prd	1.2	1.000	0.001

H_1 is the null hypothesis that the relationship between the change in mRNA expression and the binding level of transcription factor 'X' is monotone and increasing. H_2 is the null hypothesis that the relationship between the change in mRNA expression and the binding level of transcription factor 'X' is monotone and decreasing.

Table S2: Comparison of MSE of each model's fit to training data (wildtype *eve*2 pattern)

Gauss.	Gauss. + X	Gauss. *de novo*	Exp.	Exp. + X	Exp. *de novo*	Hill	Hill + X	Hill *de novo*
26.74	25.47	39.75	34.64	31.55	54.56	25.55	24.47	39.02

Gauss. represents the Gaussian model. Exp. represents the Exponential model. Model + X represents the corresponding model with a general ubiquitous activator X added to the set of transcription factors. Model *de novo* represents the corresponding model with zero mRNA expression at an additional stage 5:5-8. MSE values computed using data at last 3 time points for all models, i.e. stages 5:26-50, 5:51-75 and 5:76-100.

Table S3: Comparison of transcription factor ranking as given by ChIP-chip scores and that predicted by each model

Rank	ChIP-chip	Gauss.	Gauss. + X	Gauss. *de novo*	Exp.	Exp. + X	Exp. *de novo*	Hill	Hill + X	Hill *de novo*
1	Kr	Kr	Kr	Kr	Gt	Gt	Gt	Gt	Gt	Gt
2	Gt	Gt	Gt	Gt	Twi	Kni	Bcd	Hb	Hb	Hb
3	Kni	Hb	Hb	Hb	Kni	Hb	Kr	Sna	Cad	D
4	Hkb	D	D	D	Kr	Twi	Hb	Kni	D	Bcd
5	Cad	Kni	Kni	Hkb	Hb	Kr	D	D	Kni	Kr
6	D	Cad	Cad	Kni	Dl	Bcd	Kni	Prd	Kr	Dl
7	Tll	Hkb	Hkb	Cad	Bcd	Dl	Dl	Kr	Bcd	Run
8	Bcd	Dl	Dl	Bcd	D	D	Hkb	Bcd	Run	Cad
9	Run	Bcd	Bcd	Run	Tll	Hkb	Run	Tll	Tll	Kni
10	Dl	Tll	Tll	Dl	Sna	Tll	Tll	Cad	Prd	Prd
11	Hb	Run	Run	Tll	Hkb	Run	Cad	Run	Hkb	Hkb
12	Twi	Twi	Twi	Sna	Cad	Sna	Sna	Dl	Dl	Tll
13	Sna	Sna	Sna	Twi	Prd	Cad	Twi	Hkb	Sna	Sna
14	Prd	Prd	Prd	Prd	Run	Prd	Prd	Twi	Twi	Twi

Gauss. represents the Gaussian model. Exp. represents the Exponential model. Model + X represents the corresponding model with a general ubiquitous activator X added to the set of transcription factors. Model *de novo* represents the corresponding model with zero mRNA expression at an additional stage 5:5-8. Transcription factors ranked higher in the list correspond to stronger regulators of *eve*2 gene.

Table S4: Comparison of transcription factor ranking as given by ChIP-chip scores and that predicted by each model

Gauss.	Gauss. + X	Gauss. *de novo*	Exp.	Exp. + X	Exp. *de novo*	Hill	Hill + X	Hill *de novo*
0.78	0.78	0.78	0.33	0.42	0.62	0.10	0.56	0.32

Gauss. represents the Gaussian model. Exp. represents the Exponential model. Model + X represents the corresponding model with a general ubiquitous activator X added to the set of transcription factors. Model *de novo* represents the corresponding model with zero mRNA expression at an additional stage 5:5-8. SRCC = 1 corresponds to perfect match between the rank list given by ChIP-chip scores and the rank list predicted by a model, thus SRCC values closer to 1 are better.

B Supplementary Figures

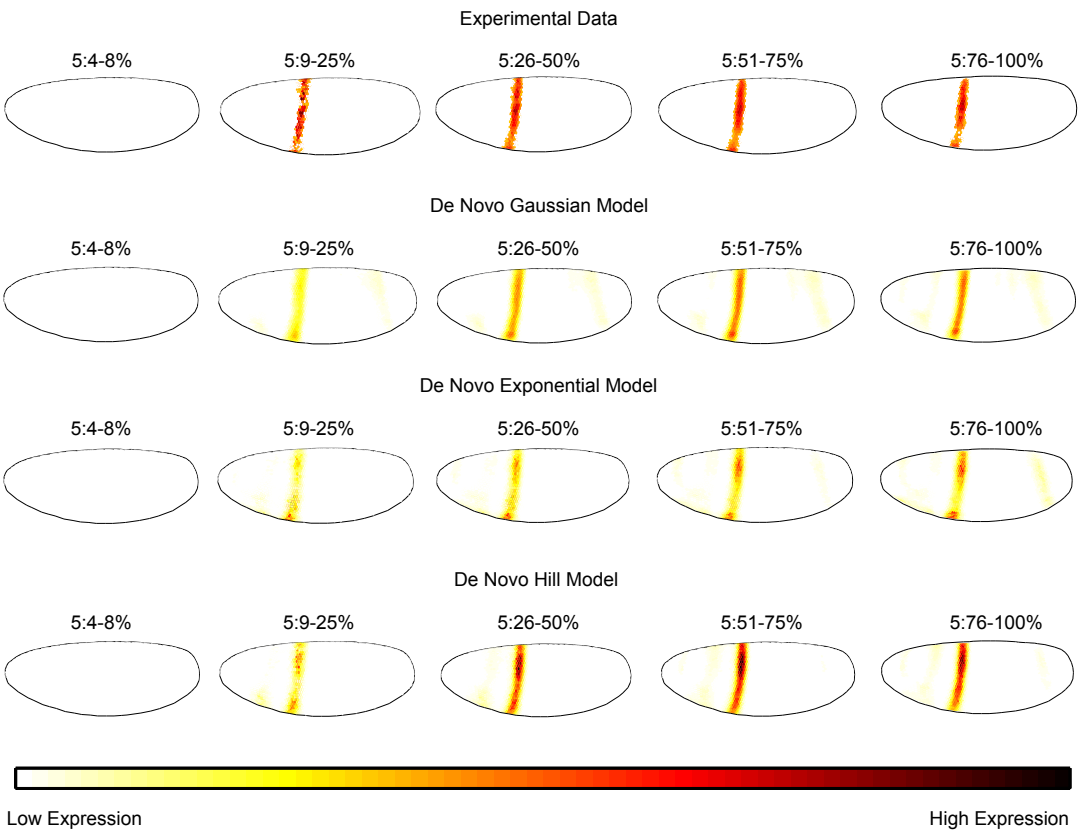

Figure S1: Predictions for training data by *de novo* models.

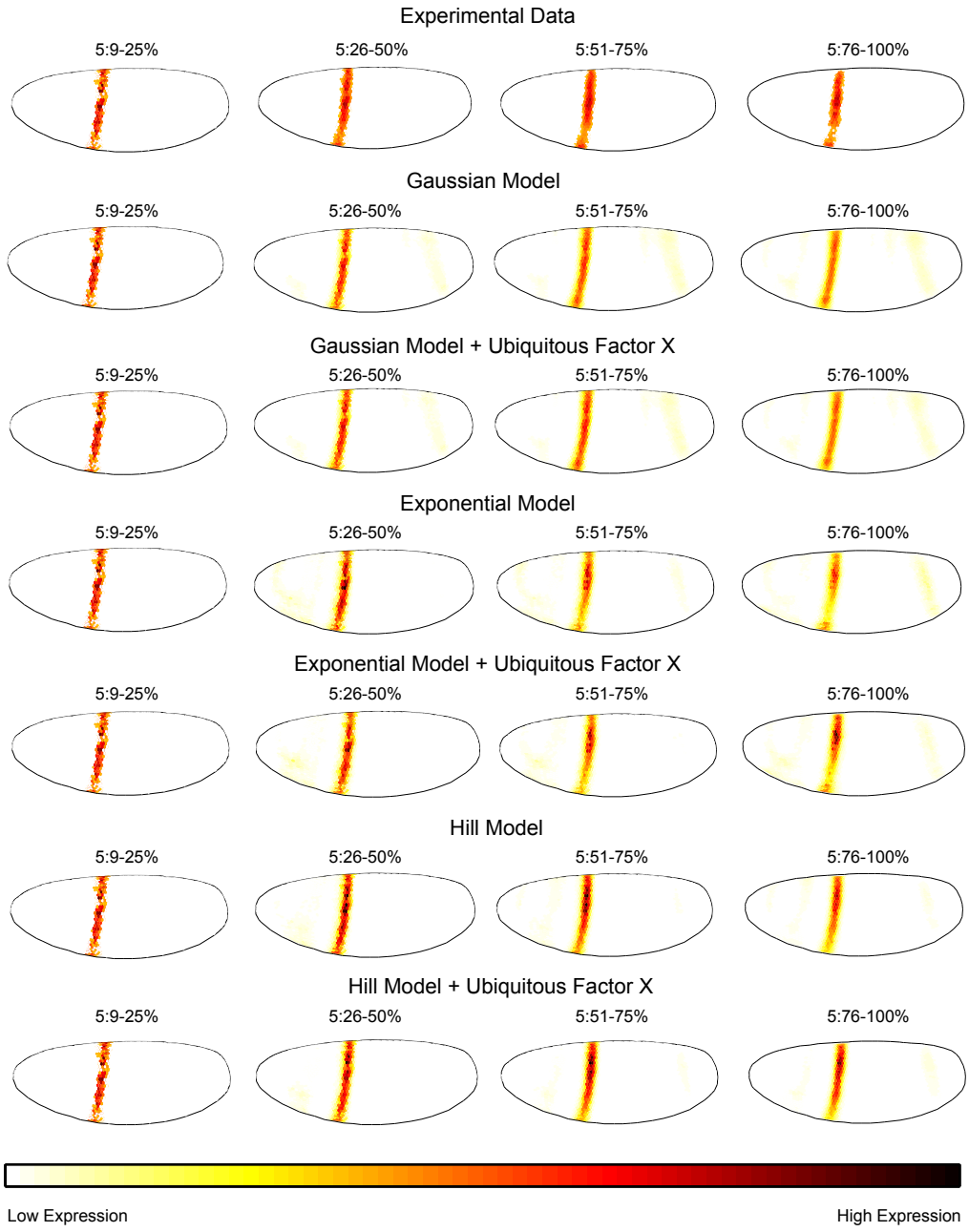

Figure S2: Predictions for training data by standard models and models with ubiquitous activator X.

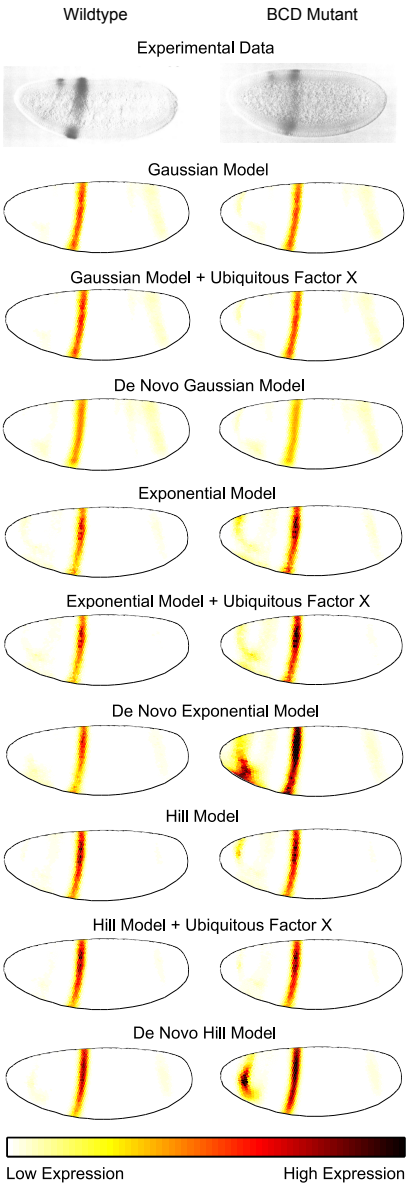

Figure S3: Comparison of model predictions for Bcd mutants.

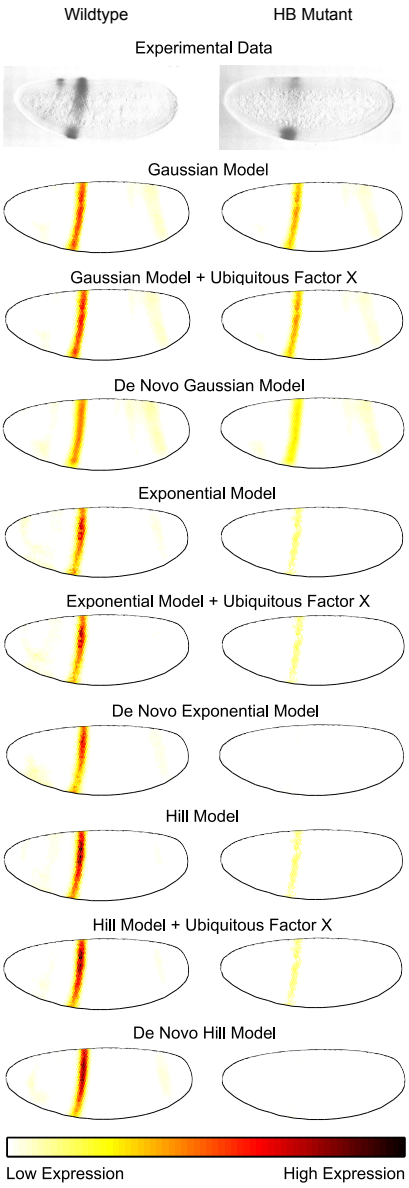

Figure S4: Comparison of model predictions for Hb mutants.

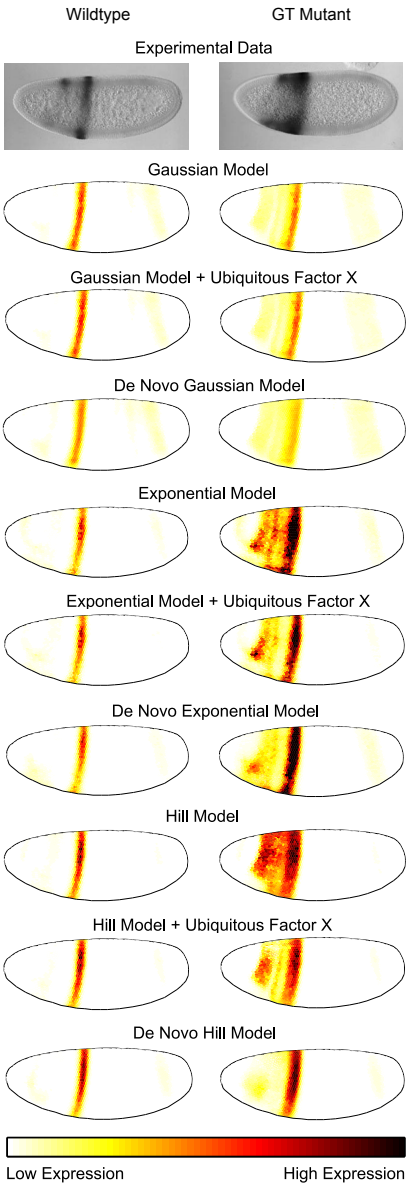

Figure S5: Comparison of model predictions for Gt mutants.

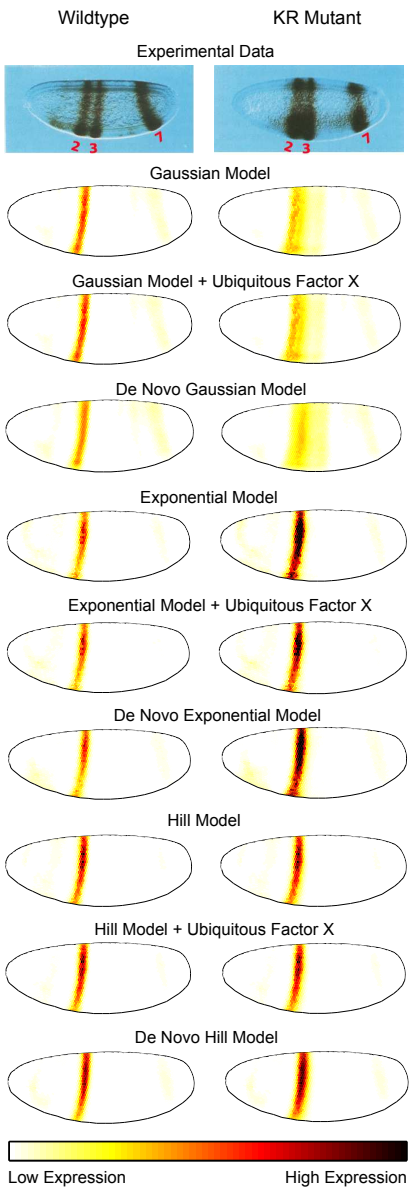

Figure S6: Comparison of model predictions for Kr mutants.

References

1. Fowlkes C, et al. (2008) A quantitative spatiotemporal atlas of gene expression in the Drosophila blastoderm. *Cell* 133(2):364-374.

2. Ilsley GR, Fisher J, Apweiler R, Depace AH, Luscombe NM (2013) Cellular resolution models for even skipped regulation in the entire Drosophila embryo. *Elife* 2:E00522.

3. Janssens H, et al. (2006) Quantitative and predictive model of transcriptional control of the Drosophila melanogaster even skipped gene. *Nat Genet* 38(10):1159-1165.

4. Kaplan T, et al. (2011) Quantitative models of the mechanisms that control genome-wide patterns of transcription factor binding during early Drosophila development. *PLoS Genet* 7(2):e1001290.

5. Aswani A, et al. (2010) Nonparametric identification of regulatory interactions from spatial and temporal gene expression data. *BMC Bioinformatics* 11:413.

6. Aswani A, Guturu H, Tomlin C (2009) System identification of Hunchback protein patterning in early Drosophila embryogenesis. *Proc IEEE Conf Decis Control* 7723-7728.

7. Biggin MD (2011) Animal transcription networks as highly connected, quantitative continua. *Dev Cell* 21(4):611-626.

8. Pique-Regi R, et al (2011) Accurate inference of transcription factor binding from DNA sequence and chromatin accessibility data. *Genome Res* 21(3):447-455.

9. De Jong H (2002) Modeling and simulation of genetic regulatory systems: a literature review. *J Comput Biol* 9(1):67-103.

10. Angeli D, Sontag ED (2003) Monotone control systems. *IEEE Trans Automat Contr* 48(10):1684-1698.

11. Venturelli OS, El-Samad H, Murray RM (2012) Synergistic dual positive feedback loops established by molecular sequestration generate robust bimodal response. *Proc Natl Acad Sci U S A* 109(48):E3324-E3333.

12. Aracena J (2008) Maximum number of fixed points in regulatory Boolean networks. *Bull Math Biol* 70(5):1398-409.

13. Cooper N, Belta C, Julius AA (2011) Genetic regulatory network identification using multivariate monotone functions. *Proc IEEE Conf Decis Control* 2208-2213.

14. Yang HT, Hsu CP, Hwang MJ (2007) An analytical rate expression for the kinetics of gene transcription mediated by dimeric transcription factors. *J Biochem* 142(2):135-144.

15. Hsia J, Holtz WJ, Huang DC, Arcak M, Maharbiz MM (2012) A feedback quenched oscillator produces Turing patterning with one diffuser. *PLoS Comput Biol* 8(1):E1002331.

16. Sauer F, Jäckle H (1991) Concentration-dependent transcriptional activation or repression by Krppel from a single binding site. *Nature* 353(6344):563-566.

17. Hülskamp M, Pfeifle C, Tautz D (1990) A morphogenetic gradient of hunchback protein organizes the expression of the gap genes Krppel and knirps in the early Drosophila embryo. *Nature* 346(6284):577-580.

18. Schulz C, Tautz D (1994) Autonomous concentration-dependent activation and repression of Krppel by hunchback in the Drosophila embryo. *Development* 120(10):3043-3049.

19. Zuo P, et al. (1991) Activation and repression of transcription by the gap proteins hunchback and Krppel in cultured Drosophila cells. *Genes Dev* 5:254-264.

20. Kelley KM, Wang H, Ratnam M (2003) Dual regulation of ets-activated gene expression by SP1. *Gene* 307:87-97.

21. Bieler J, Pozzorini C, Naef F (2011) Whole-embryo modeling of early segmentation in Drosophila identifies robust and fragile expression domains. *Biophys J* 101(2):287-296.

22. Papatsenko D, Levine MS (2008) Dual regulation by the Hunchback gradient in the Drosophila embryo. *Proc Natl Acad Sci U S A* 105(8):2901-2906.

23. MacArthur S, et al. (2009) Developmental roles of 21 Drosophila transcription factors are determined by quantitative differences in binding to an overlapping set of thousands of genomic regions. *Genome Biol* 10(7):R80.

24. Small S, Blair A, Levine M (1996) Regulation of two pair-rule stripes by a single enhancer in the Drosophila embryo. *Dev Biol* 175(2):314-324.

25. Small S, Blair A, Levine M (1992) Regulation of even-skipped stripe 2 in the Drosophila embryo. *EMBO J* 11(11):4047-4057.

26. Small S, Kraut R, Hoey T, Warrior R, Levine M (1991) Transcriptional regulation of a pair-rule stripe in Drosophila. *Genes Dev* 5(5):827-839.

27. Luengo Hedriks CL, et al. (2006) Three-dimensional morphology and gene expression in the Drosophila blastoderm at cellular resolution I: data acquisition pipeline. *Genome Biol* 7(12):R123.

28. Hall P, Heckman NE (2000) Testing for monotonicity of a regression mean by calibrating for linear functions. *Ann Stat* 28(1):20-39.

29. Fujioka M, Emi-Sarker Y, Yusibova G, Goto T, Jaynes J (1999) Analysis of an even-skipped rescue transgene reveals both composite and discrete neuronal and early blastoderm enhancers, and multi-stripe positioning by gap gene repressor gradients. *Development* 126(11):2527-2538.

30. Arnosti D, Barolo S, Levine M, Small S (1996) The *eve* stripe 2 enhancer employs multiple modes of transcriptional synergy. *Development* 122(1):205-214.

31. Segal E, Raveh-Sadka T, Schroeder M, Unnerstall U, Gaul U (2008) Predicting expression patterns from regulatory sequence in Drosophila segmentation. *Nature* 451(7178):535-540.

32. Fakhouri W, et al. (2010) Deciphering a transcriptional regulatory code: modeling short-range repression in the Drosophila embryo. *Mol Syst Biol* 6:341.

33. Li XY, et al. (2008) Transcription factors bind thousands of active and inactive regions in the Drosophila blastoderm. *PLoS Biol* 6(2):e27.

34. Gregor T, Tank DW, Wieschaus EF, Bialek W (2007) Probing the limits to positional information. *Cell* 130(1):153-164.

35. Kazemian M, et al. (2010) Quantitative analysis of the Drosophila segmentation regulatory network using pattern generating potentials. *PLoS Biol* 8(8):e1000456.

36. He F, et al. (2008) Probing intrinsic properties of a robust morphogen gradient in Drosophila. *Dev Cell* 15(4):558-567.

37. Gregor T, Wieschaus EF, McGregor AP, Bialek W, Tank DW (2007) Stability and nuclear dynamics of the bicoid morphogen gradient. *Cell* 130(1):141-152.

38. Crauk O, Dostatni N (2005) Bicoid determines sharp and precise target gene expression in the Drosophila embryo. *Curr Biol* 15(21):1888-1898.

39. Licht JD, Grossel MJ, Figge J, Hansen UM (1990) Drosophila Krppel protein is a transcriptional repressor. *Nature* 346(6270):76-79.

40. Stanojevic D, Small S, Levine M (1991) Regulation of a segmentation stripe by overlapping activators and repressors in the Drosophila embryo. *Science* 254(5036):1385-1387.

41. Goto T, Macdonald P, Maniatis T (1989) Early and late periodic patterns of even skipped expression are controlled by distinct regulatory elements that respond to different spatial cues. *Cell* 57(3):413-422.

42. Struhl G, Struhl K, Macdonald PM (1989) The gradient morphogen bicoid is a concentration-dependent transcriptional activator. *Cell* 57(7):1259-1273.

43. Andrioli LP, Vasisht V, Theodosopoulou E, Oberstein A, Small S (2002) Anterior repression of a Drosophila stripe enhancer requires three position-specific mechanisms. *Development* 129(21):4931-4940.

44. Löhr U, Chung HR, Beller M, Jäckle H (2009) Antagonistic action of Bicoid and repressor Capicua determines the spatial limits of Drosophila head gene expression domains. *Proc Natl Acad Sci U S A* 106(51):21695-21700.

45. Bellaïche Y, Bandyopadhyay R, Desplan C, Dostatni N (1996) Neither the homeodomain nor the activation domain of bicoid is specifically required for its down-regulation by the torse receptor tyrosine kinase cascade. *Development* 122(11):3499-3508.

46. Singh N, Zhu W, Hanes SD (2005) Sap18 is required for the maternal gene bicoid to direct anterior patterning in Drosophila melanogaster. *Dev Biol* 278(1):242-254.

47. Janody F, Sturny R, Catala F, Desplan C, Dostatni N (2000) Phosphorylation of bicoid on MAP-kinase sites: contribution to its interaction with the torso pathway. *Development* 127(2):279-289.

48. Zhao C, et al (2002) The activity of the Drosophila morphogenetic protein Bicoid is inhibited by a domain located outside its homeodomain. *Development* 129(7):1669-1680.

49. Fu D, Ma J (2005) Interplay between positive and negative activities and influence the role of Bicoid in transcription. *Nucleic Acids Res* 33(13):3985-3993.

50. Gutierrez PS, Monteoliva D, Diambra L (2012) Cooperative binding of transcription factors promotes bimodal gene expression response. *PLoS One* 7(9):e44812.

51. Biggin MD, McGinnis W (1997) Regulation of segmentation and segmental identity by Drosophila homeoproteins: the role of DNA binding in functional activity and specificity. *Development* 124:4425-4433.

52. Liang Z, Biggin MD (1998) Eve and ftz regulate a wide array of genes in blastoderm embryos: the selector homeoproteins directly or indirectly regulate most genes in Drosophila. *Development* 125:4471-4482.

53. Jenner RG, et al (2009) The transcription factors T-bet

54. Elemento O, Tavazoie S (2005) Fast and systematic genome-wide discovery of conserved regulatory elements using a non-alignment based approach. *Genome Biol* 6(2):R18.

55. Oh YM, Kim JK, Choi S, Yoo JY (2011) Identification of co-occurring transcription factor binding sites from DNA sequence using clustered position weight matrices. *Nucleic Acids Res* 40(5):e38.