Part 2: Application to the *Drosophila* segmentation network

Theodore J. Perkins

McGill University

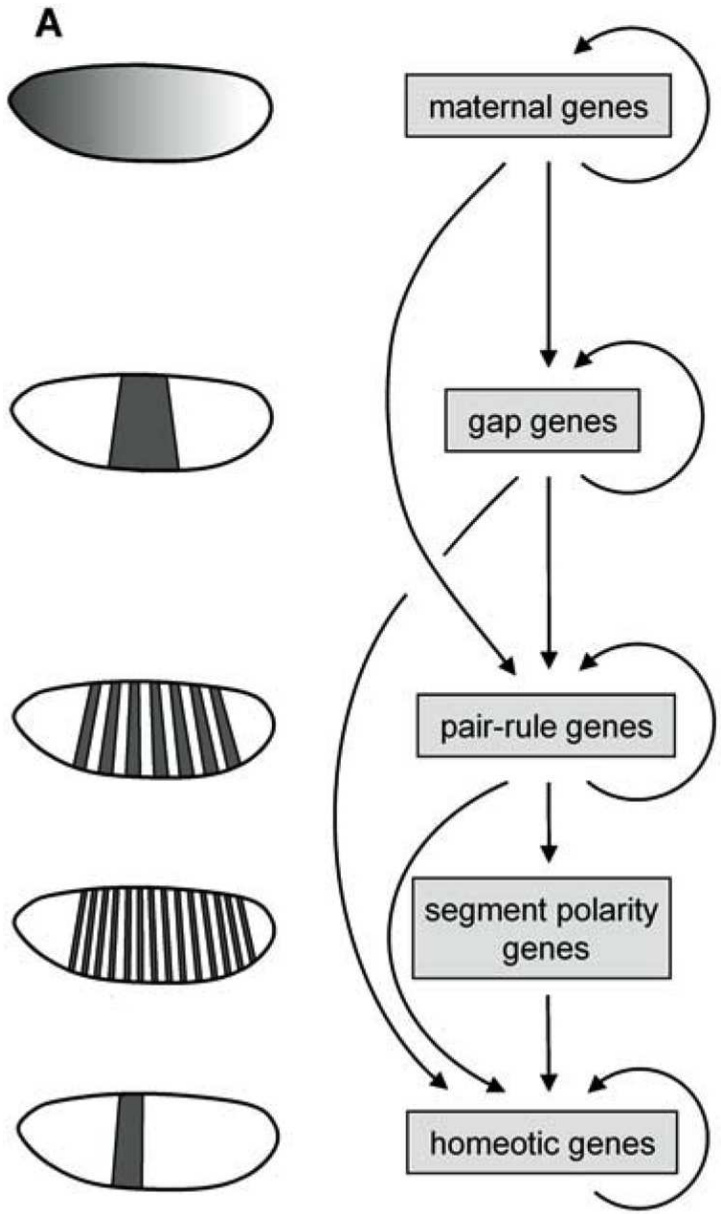
perkins@mcb.mcgill.ca www.mcb.mcgill.ca/∼perkins

Outline

- The segmentation network
- *In situ* hybridization expression data
- Modeling work of Reinitz *et al.*
  - Model formulation & fitting
  - Shifting of gap gene domains
- My own modeling work
  - Faster model-fitting & recovery of Reinitz's results
  - Alternative network architectures
  - Alternative model formulations

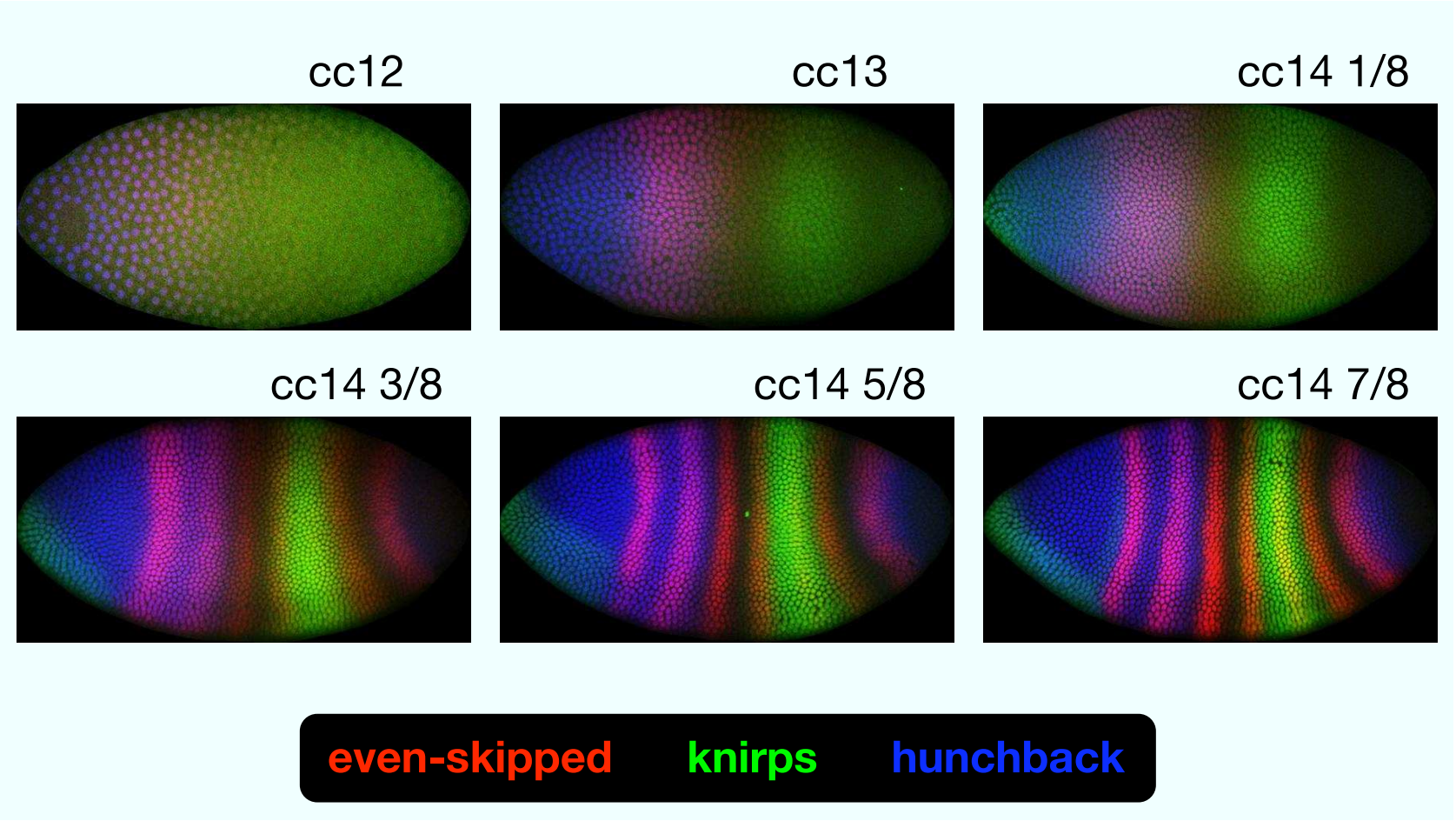
The segmentation network

- Establishes “positional information” – dividing anterior-posterior axis of embryo into body segments
- Broad architecture is a genetic cascade with groups of interacting genes
- Expression of gap and pair-rule genes is transient



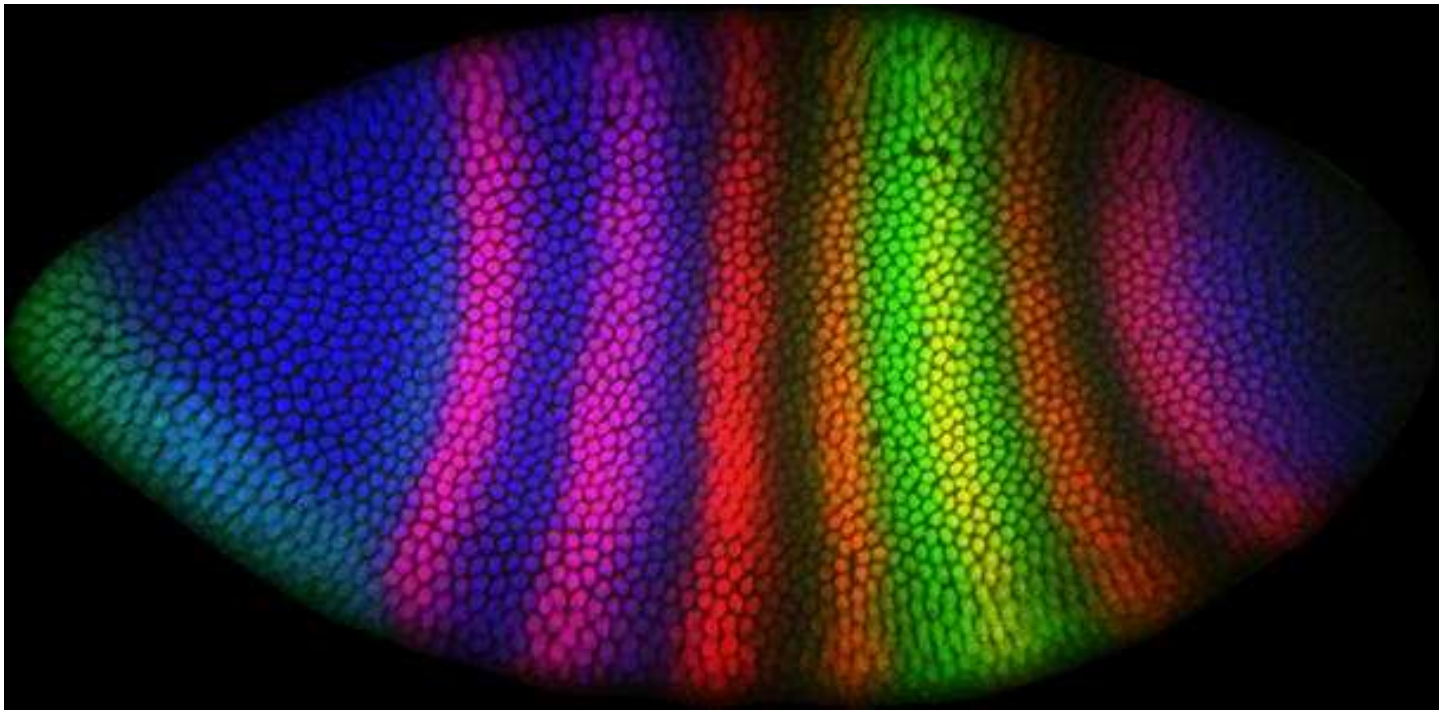
Pattern formation

Pattern form during cleavage cycles 13 (or 12?) and 14A, a period of $\sim$ 1.5 hours. (Except maternal genes, whose gradients are established much earlier.)



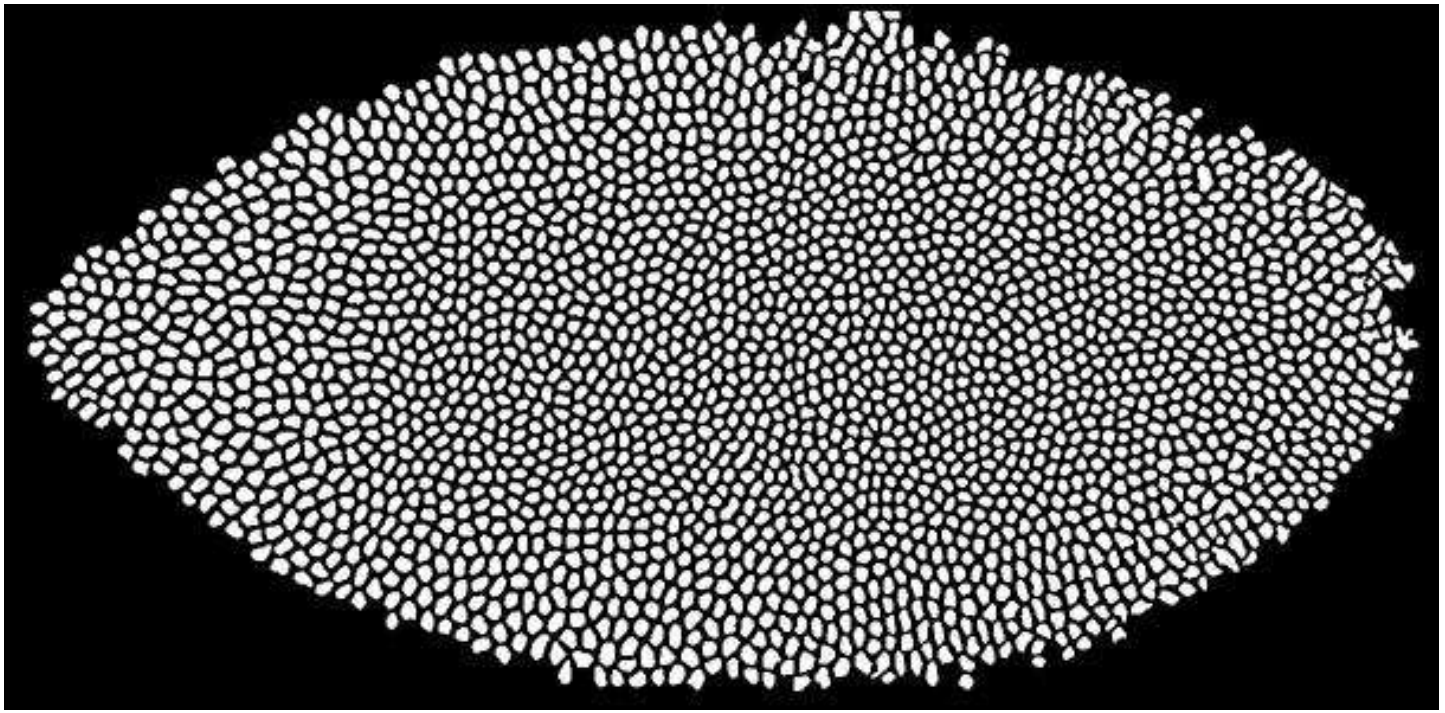
In situ hybridization data

Up to three (species of) proteins at a time are labeled with fluorescent antibodies and imaged under a microscope.



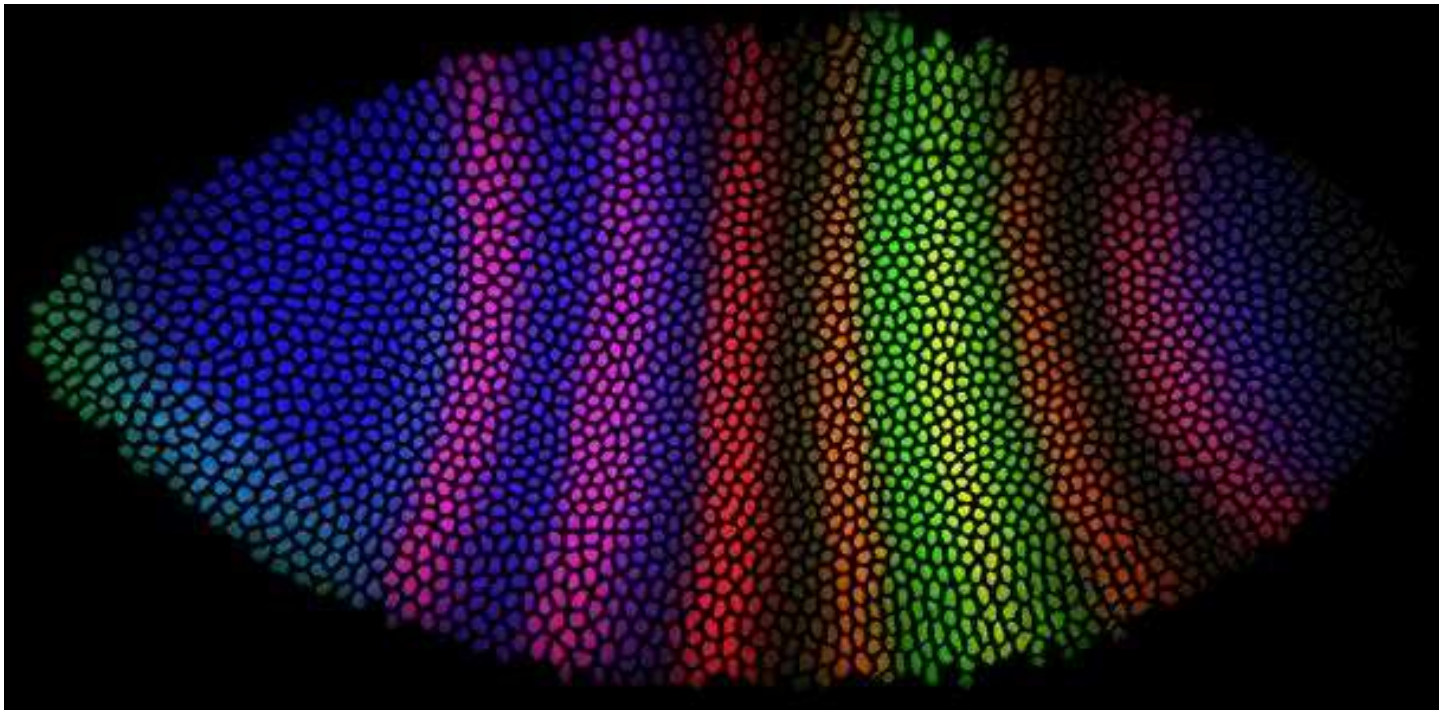
Nuclear masking

Nuclei are detected by thresholding based on intensity, and finding connected groups of pixels.



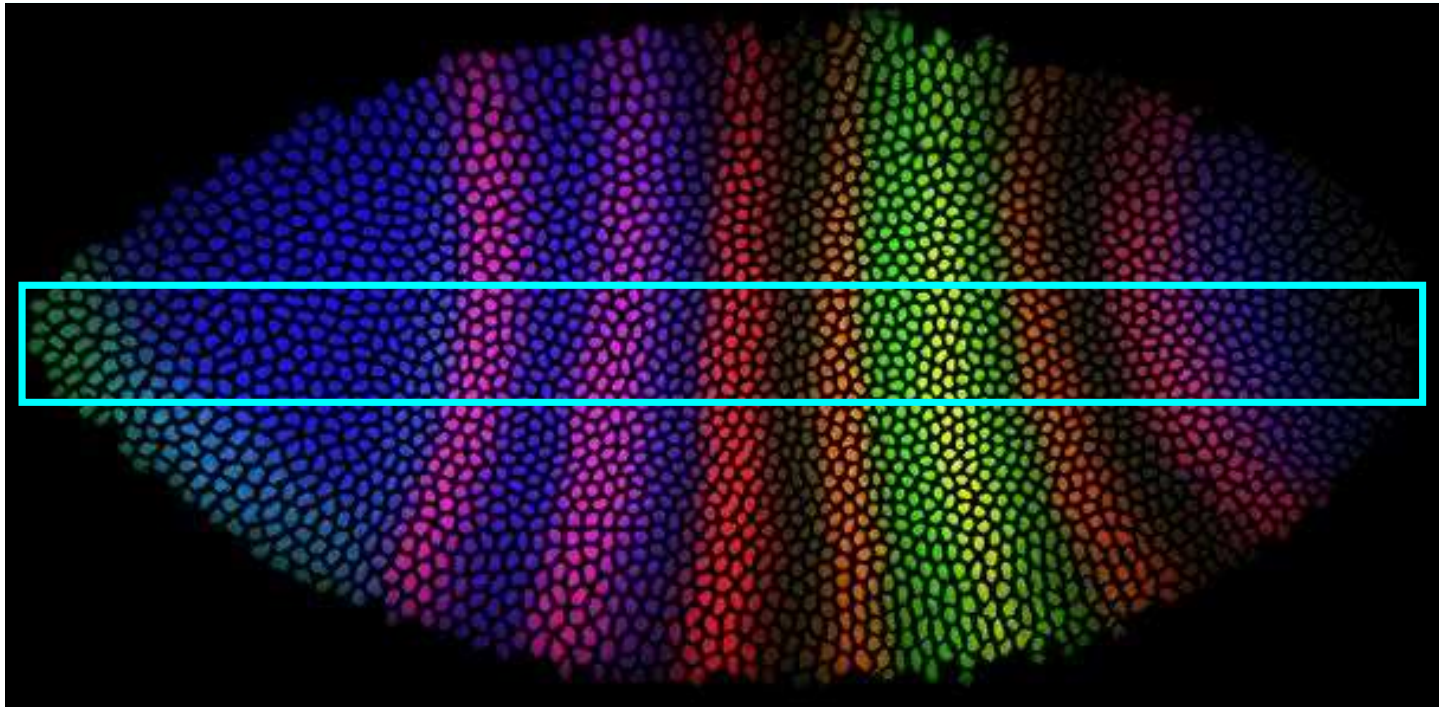
Per nucleus intensities

Intensity in each nucleus in each channel is computed by averaging over the pixels associated to the nucleus.



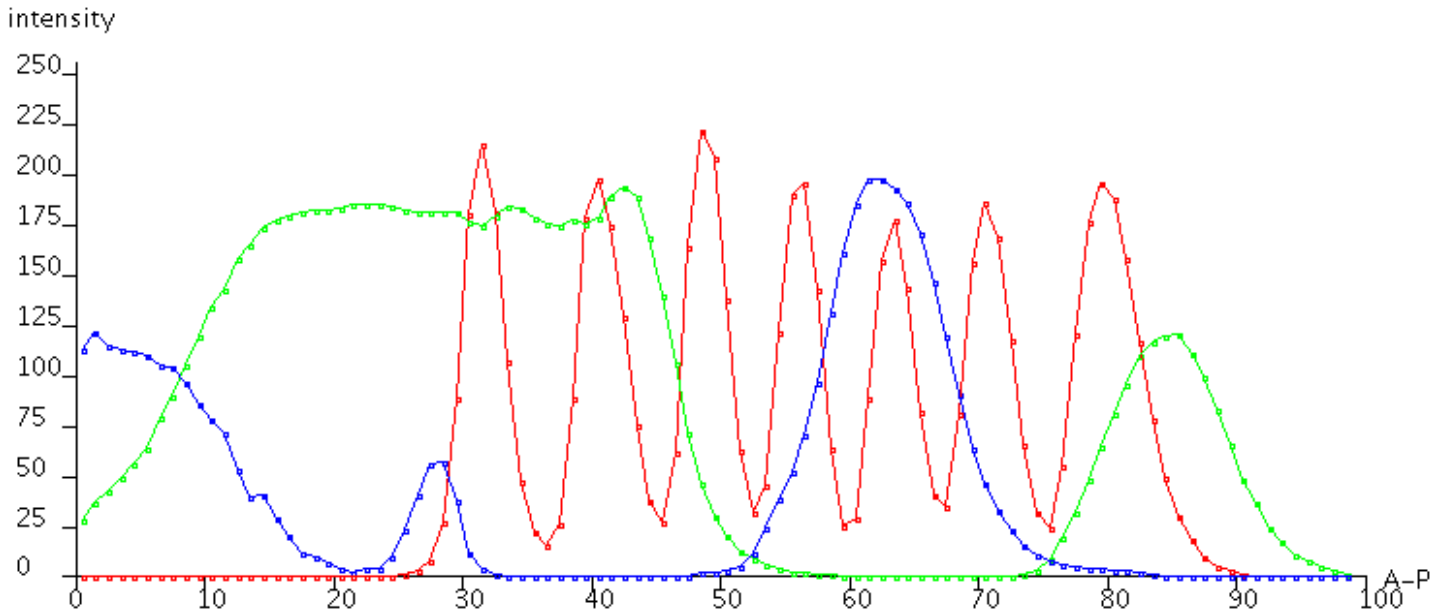
Creating expression profiles

Intensities are averaged vertically within a window centered along the A-P axis of the embryo, at a resolution of 1% embryo length (EL).



Creating expression profiles

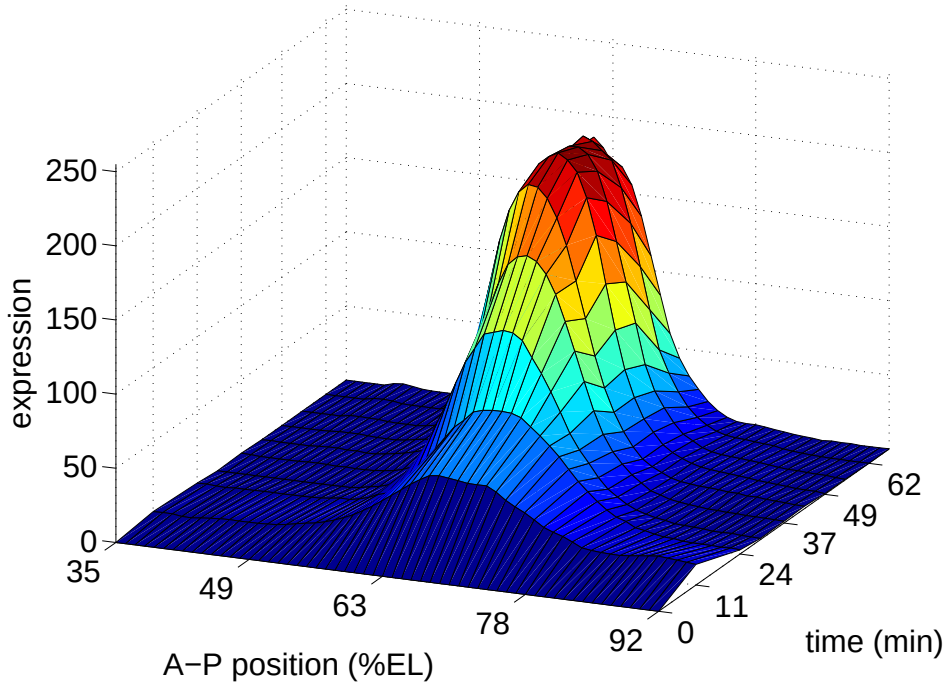
Intensities are averaged vertically within a window centered along the A-P axis of the embryo, at a resolution of 1% embryo length (EL).



eve in red, *hb* in green, *kni* in blue (green and blue are switched from previous slide)

Creating a time series

- Labeling and imaging destroys the embryo, so the preceding must be repeated – at many different times and for different combinations of genes.
- The profiles are combined to produce “average” wild-type expression at 10 times during cleavage cycles 12-14A, at 1% EL resolution (shown here between 35% EL and 92% EL).

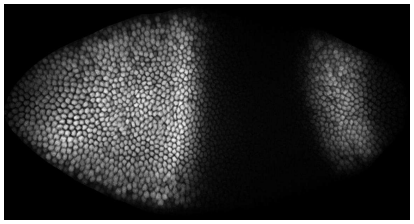


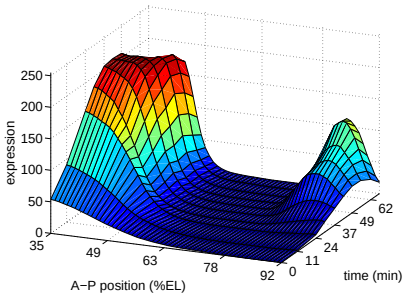
Some previous work based on this data

- Reinitz & Sharp (1995) “Mechanism of *eve* stripe formation” *Mechanisms of Development* 49: 133+
- Jaeger et al. (2004) “Dynamic control of positional information in the early *Drosophila* embryo” *Nature* 430: 368+
- Jaeger et al. (2004) “Dynamical Analysis of Regulatory Interactions in the Gap Gene System of *Drosophila melanogaster*” *Genetics* 167: 1721+
- Perkins et al. (2006) “Reverse Engineering the Gap Gene Network of *Drosophila melanogaster*” *PLoS Comp. Bio 2*(5):e51
- Janssens et al. (2006) “Quantitative and predictive model of transcriptional control of the *Drosophila melanogaster even skipped* gene” *Nature Genetics* 38: 1159+

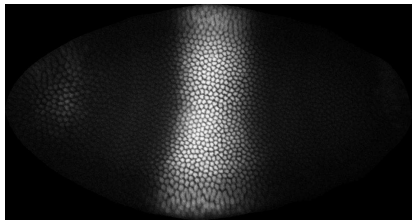
We'll consider models of pattern formation for the four trunk gap genes (hb, Kr, kni, gt) with exogenous inputs (bcd,cad,tll):

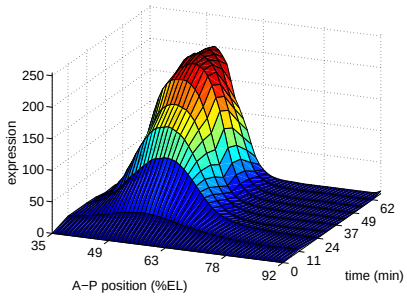
hb



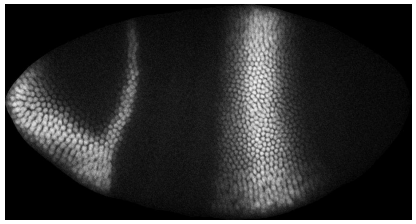


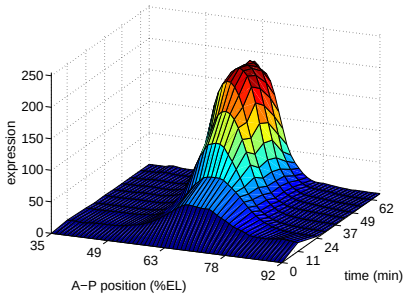
Kr



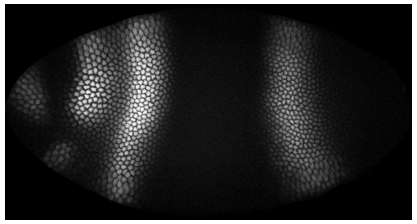


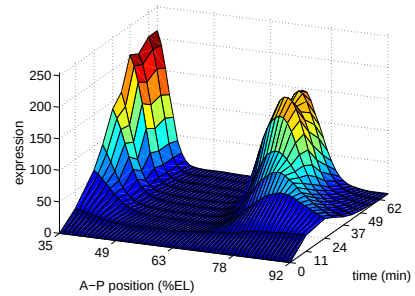
kni



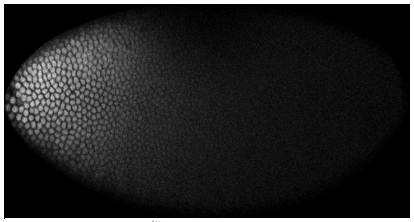


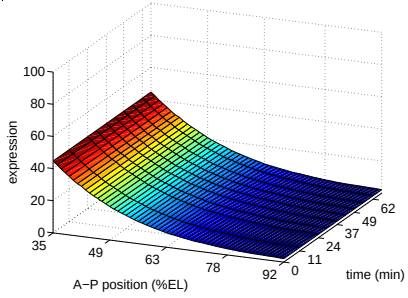
gt



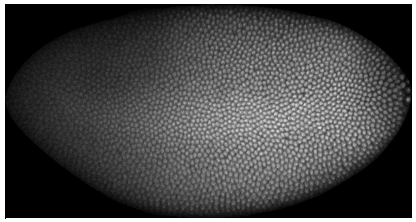


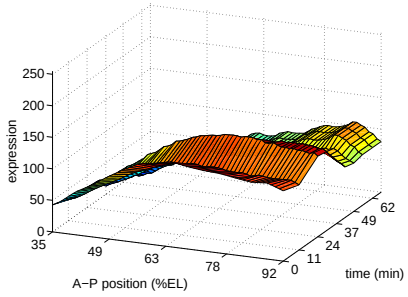
bcd



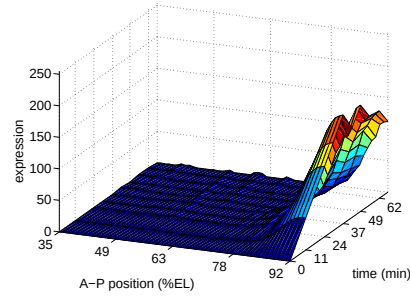


cad





tll



A PDE model for protein levels

Let x denote space, t time, and $v^a(x,t)$ the expression of protein a at space x and time t. Then:

$$\frac{\partial v^a(x,t)}{\partial t} = \underbrace{P^a(v(x,t))}_{\text{production}} - \underbrace{\gamma^a v^a(x,t)}_{\text{decay}} + \underbrace{D^a \frac{\partial^2 v^a(x,t)}{\partial x^2}}_{\text{diffusion}}$$

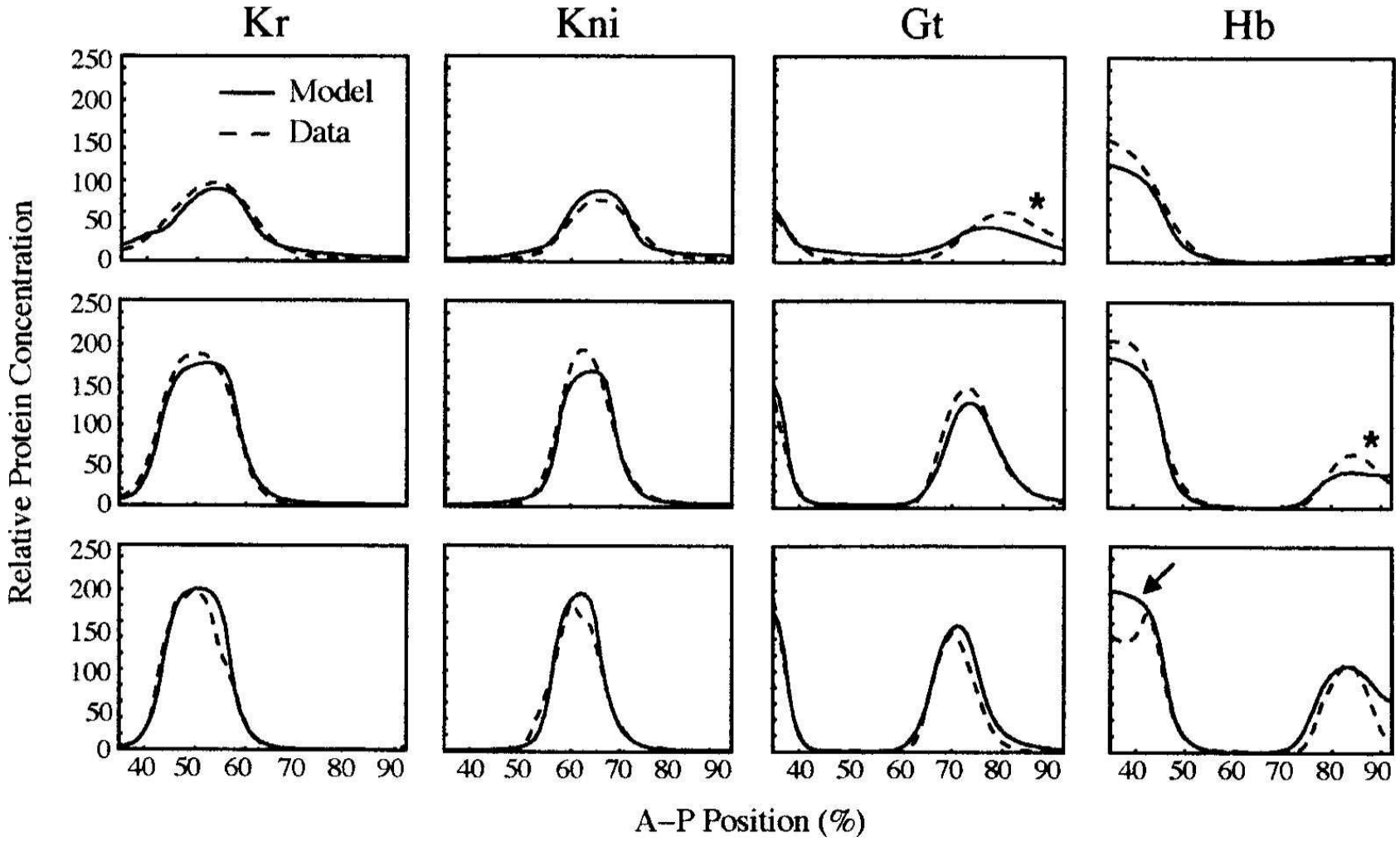
where

$$P^a(v(x,t)) = R^a g\left(\sum_b T^{ab} v^b(x,t) + h^a\right)$$

where $g(u) = \frac{1}{2}\left(\frac{u}{\sqrt{u^2+1}} + 1\right)$ is sigmoidal in the range $[0,1]$.

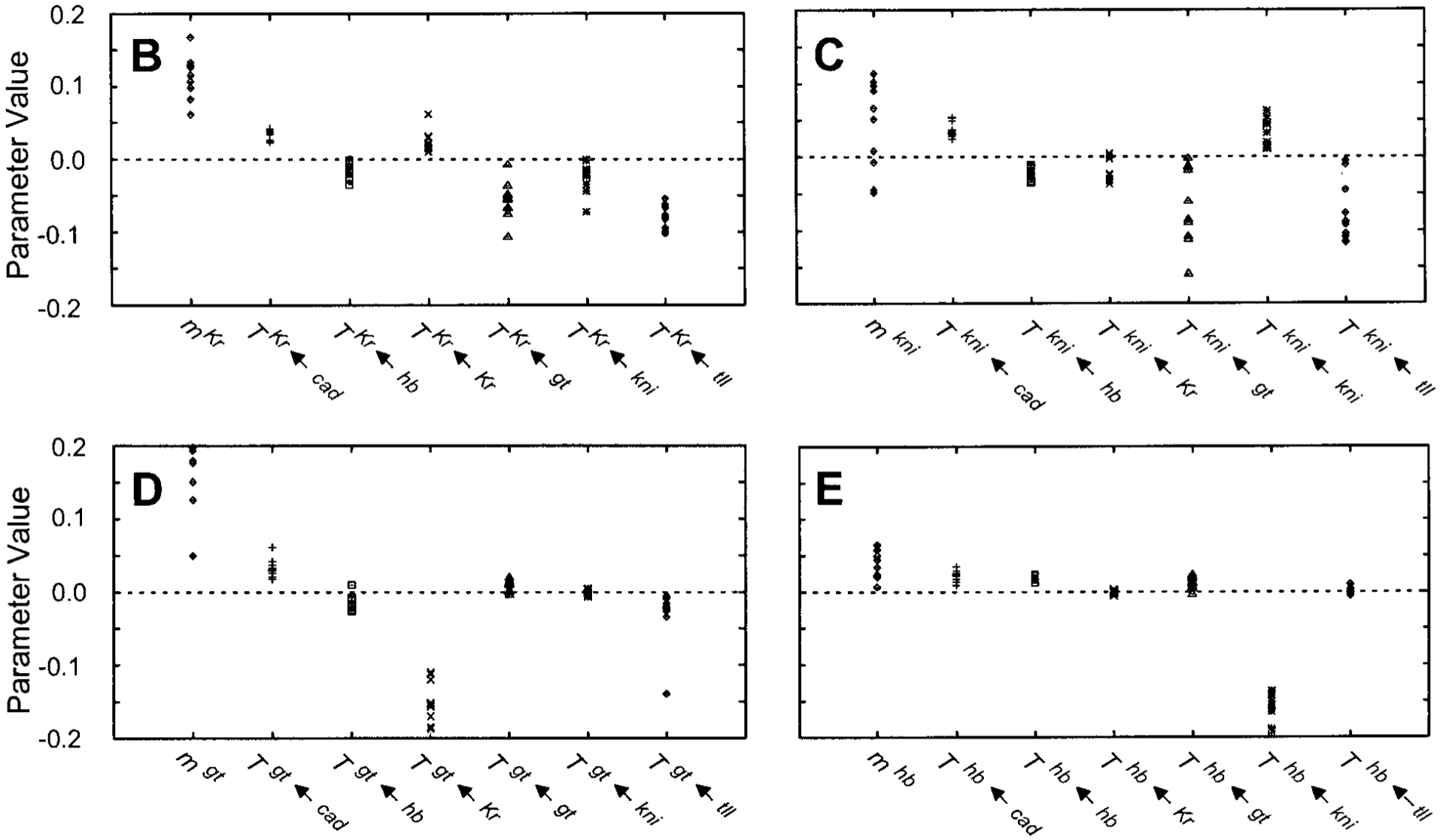
Can the model explain the data?

They fit all parameters using multiple runs of a parallel simulated annealing algorithm. Simulated expression from the best fit:



How do the fitted models explain the data?

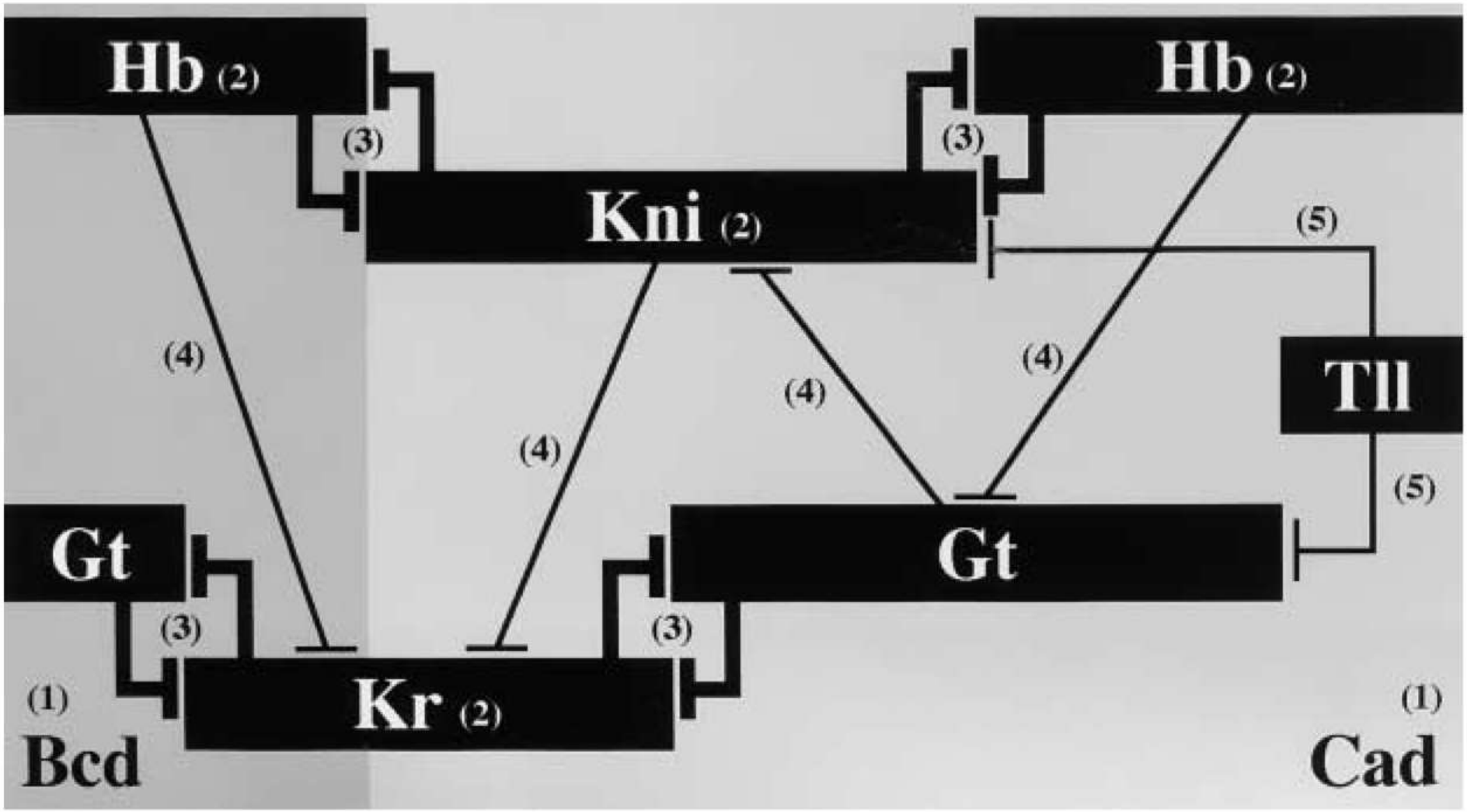
Variability in fitted parameters in best 10 runs:



Signs of fitted parameters in best 10 runs

A	*bcd*	*cad*	*hb*	*Kr*	*gt*	*kni*	*tll*
hb	0/0/10	0/0/10	0/0/10	*0/10/0*	0/2/8	**10/0/0**	*1/8/1*
Kr	0/0/10	0/0/10	**9/1/0**	0/0/10	**10/0/0**	**9/1/0**	**10/0/0**
gt	0/0/10	0/0/10	**9/0/1**	**10/0/0**	0/3/7	*1/9/0*	**10/0/0**
kni	3/0/7	0/0/10	**10/0/0**	**6/4/0**	**9/1/0**	0/0/10	**10/0/0**

Estimate of regulatory architecture



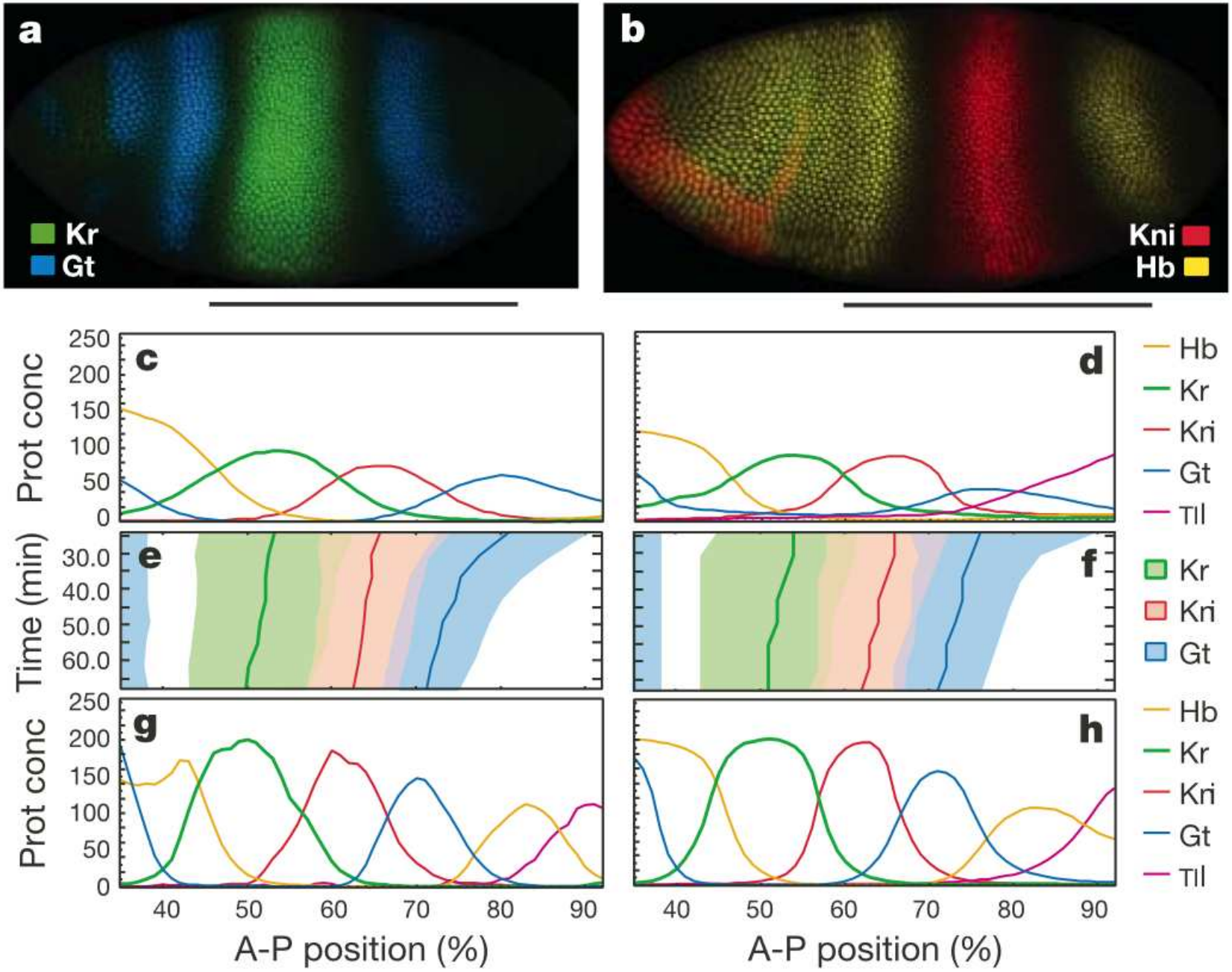
Comparison to previous models

A

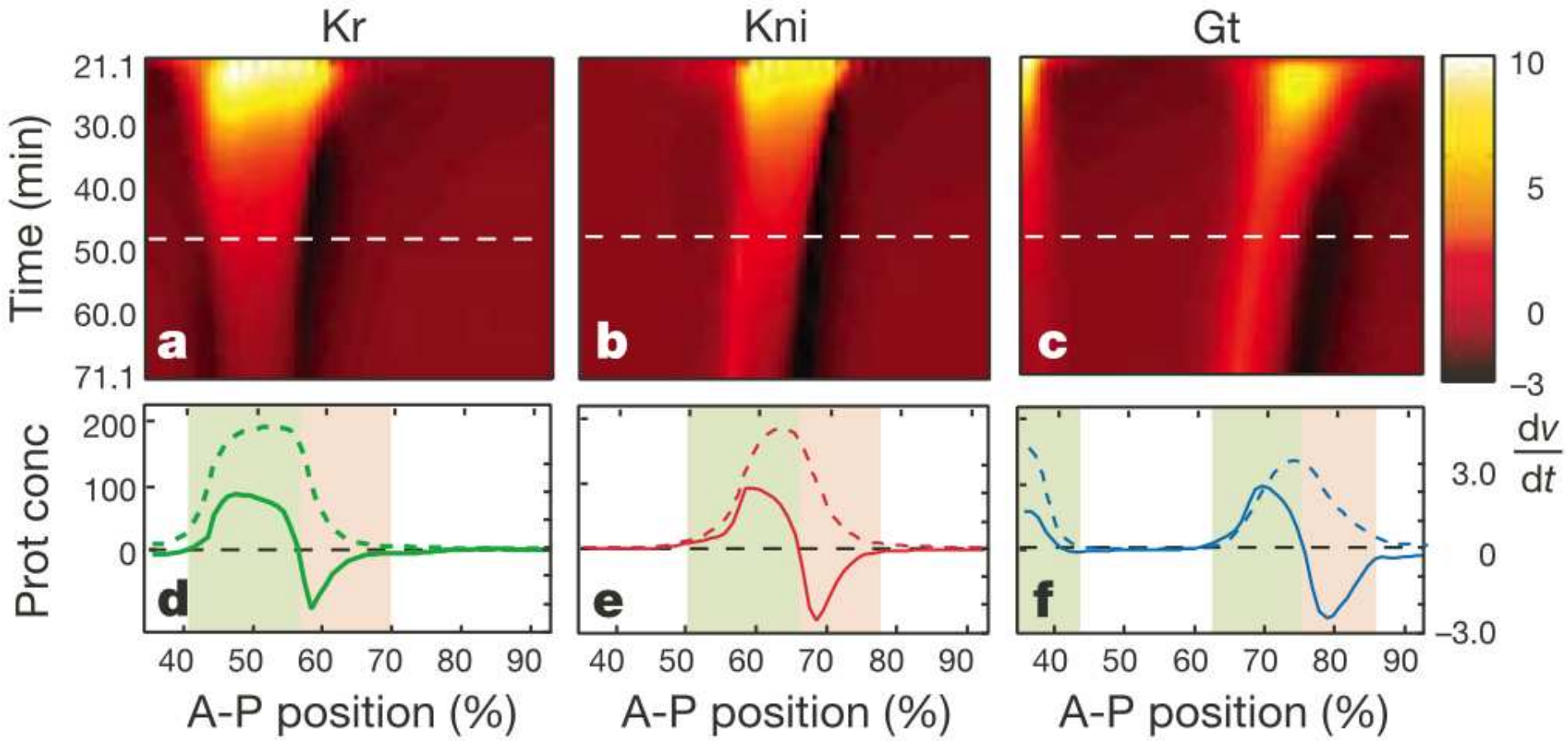
Network	Hb regulation							Kr regulation							Gt regulation							Kni regulation						
	Bcd	Cad	Hb	Kr	Gt	Kni	Tll	Bcd	Cad	Hb	Kr	Gt	Kni	Tll	Bcd	Cad	Hb	Kr	Gt	Kni	Tll	Bcd	Cad	Hb	Kr	Gt	Kni	Tll
R-P & J	+	·	+	−	·	·	·	+	·	+	−	·	−	−	−	+	+	·	−	·	−	−	+	+	−	+	−	·	−
S & T	+	·	+	−	·	·	·	+	·	+	−	·	−	−	·	+	+	−	−	·	·	·	+	+	−	·	−	·	·
Jaeger et al.	+	+	+	0	+	−	0	+	+	−	+	−	−	−	+	+	−	−	+	0	−	+	+	−	−	−	+	−

- R-P & J from: Rivera-Pomar & Jäckle (1996) Trends in Genetics 12:478–483
- S & T from: Sanchez and Thieffry (2001) Journal of Theoretical Biology 211: 115–151
- Jaeger et al. from: Jaeger et al. (2004) Nature 430:369–371, and Jaeger et al. (2004) Genetics 167:1721–1737

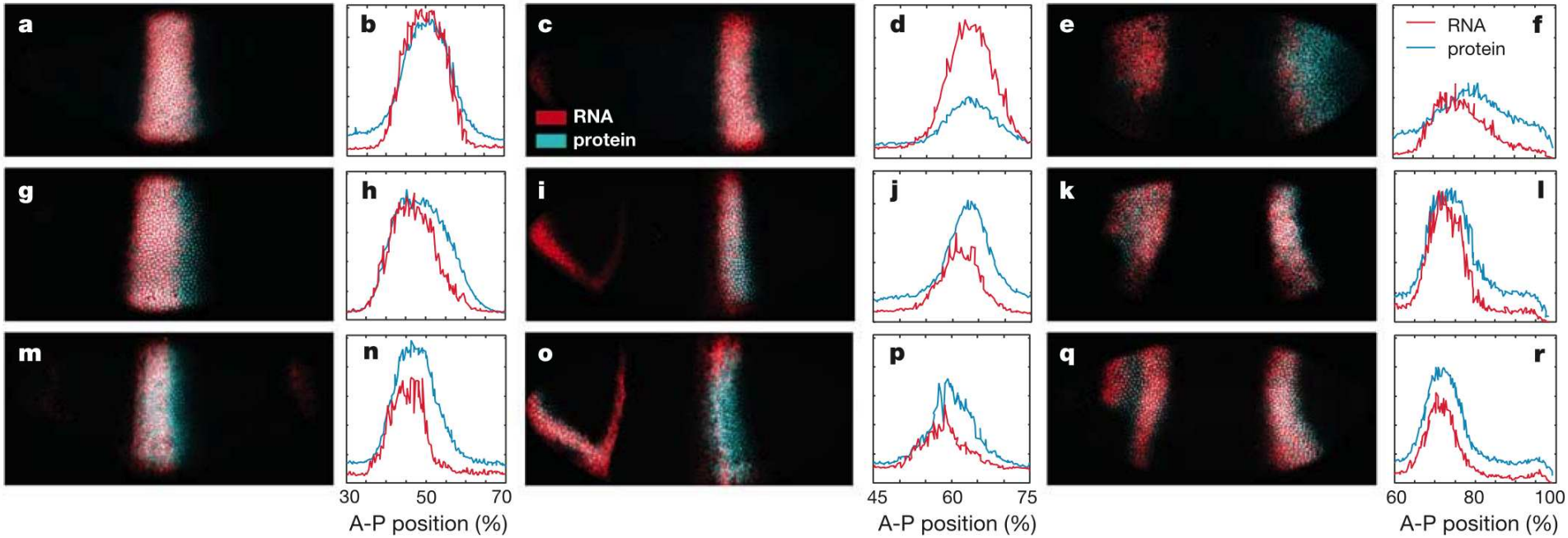
Shifting of Kr, Kni, and post-Gt domains



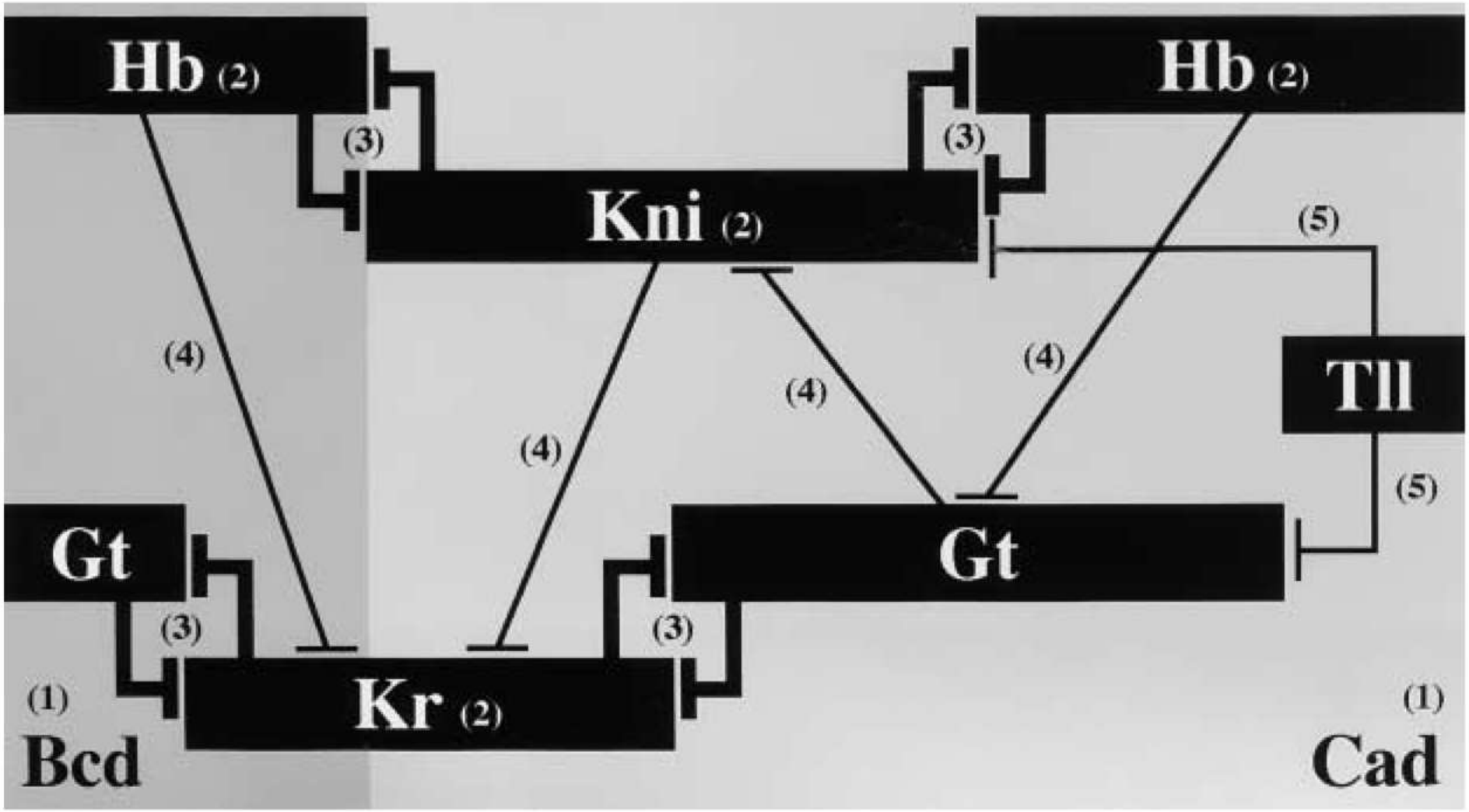
Model prediction of production rate



Verification by joint mRNA and protein staining



Repressive chain causing the shift: Hb-Gt-Kni-Kr



Limitations of the Jaeger et al. methodology

- Computation time . . .

Limitations of the Jaeger et al. methodology

- Computation time of 2 CPU years
- No explicit testing of alternative regulatory hypotheses/architectures
- No testing of alternative modeling formalisms

Perkins et al. (2006)

- First goal: Speed up model fitting!
- Second goal: Explore alternative network architectures and modeling formalisms.

First fit

For first experiment, assume same model type as Jaeger et al.:

$$\frac{\partial v^a(x,t)}{\partial t} = \underbrace{P^a(v(x,t))}_{\text{production}} - \underbrace{\gamma^a v^a(x,t)}_{\text{decay}} + \underbrace{D^a \frac{\partial^2 v^a(x,t)}{\partial x^2}}_{\text{diffusion}}$$

where

$$P^a(v(x,t)) = R^a g\left(\sum_b T^{ab} v^b(x,t) + h^a\right)$$

where $g(u) = \frac{1}{2}\left(\frac{u}{\sqrt{u^2+1}} + 1\right)$.

How to fit faster, without sacrificing quality?

1. Estimate $\frac{\partial v^a(x,t)}{\partial t}$

   (More precisely, estimate $P^a(x,t)$, γ^a, D^a)

2. Fit estimated production rates – essentially a logistic regression problem

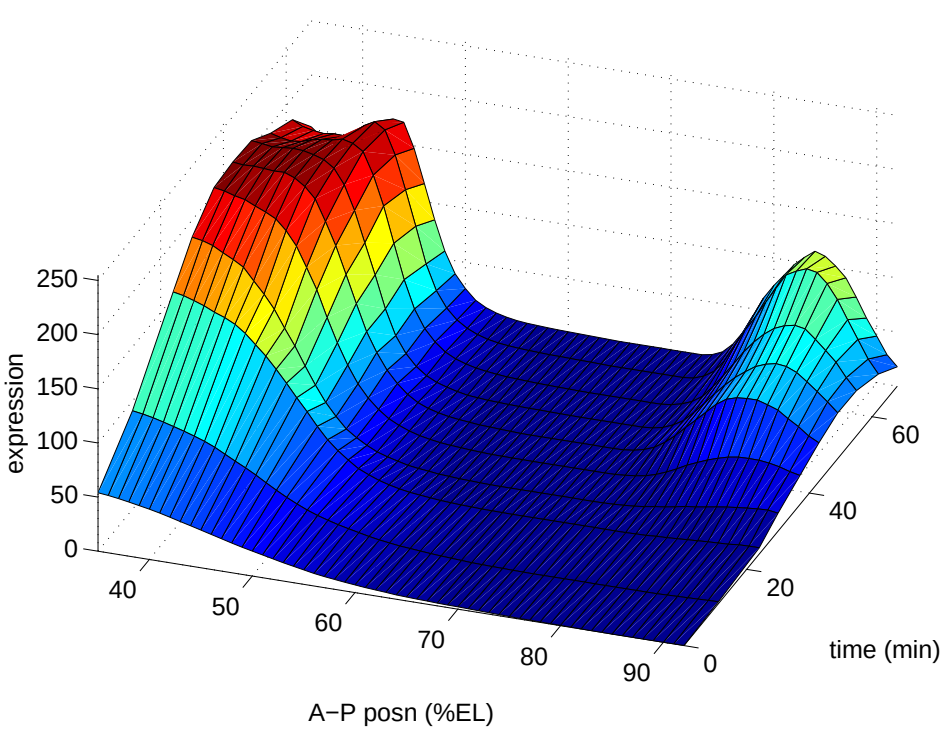
   (Optimize R^a, T^{ab}, h^a so that $P^a(x,t) \approx R^a g\left(\sum_b T^{ab} v^b(x,t) + h^a\right)$)

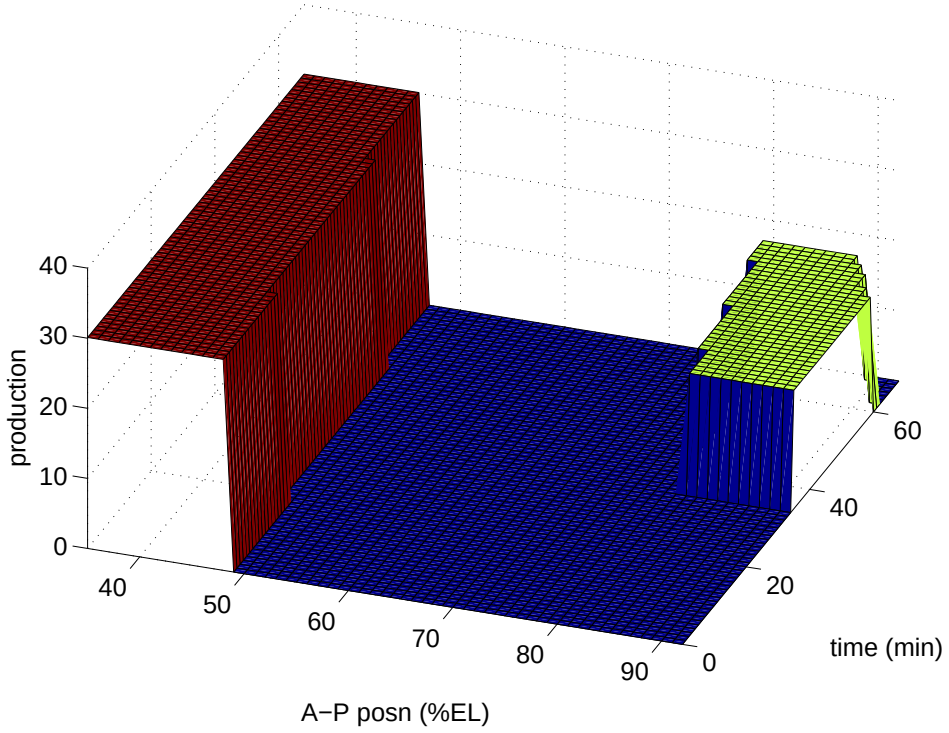
3. Starting from $R^a, T^{ab}, h^a, \gamma^a, D^a$ above, tune parameters so that simulated expression fits data well

Step 1: Estimate $P^a(x,t), \gamma^a, D^a$

$$\frac{\partial v^a(x,t)}{\partial t} = P^a(x,t) - \gamma^a v^a(x,t) + D^a \frac{\partial^2 v^a(x,t)}{\partial x^2}$$

- Production given by quadrilateral patches of space-time
- Optimize so simulated expression matches observed

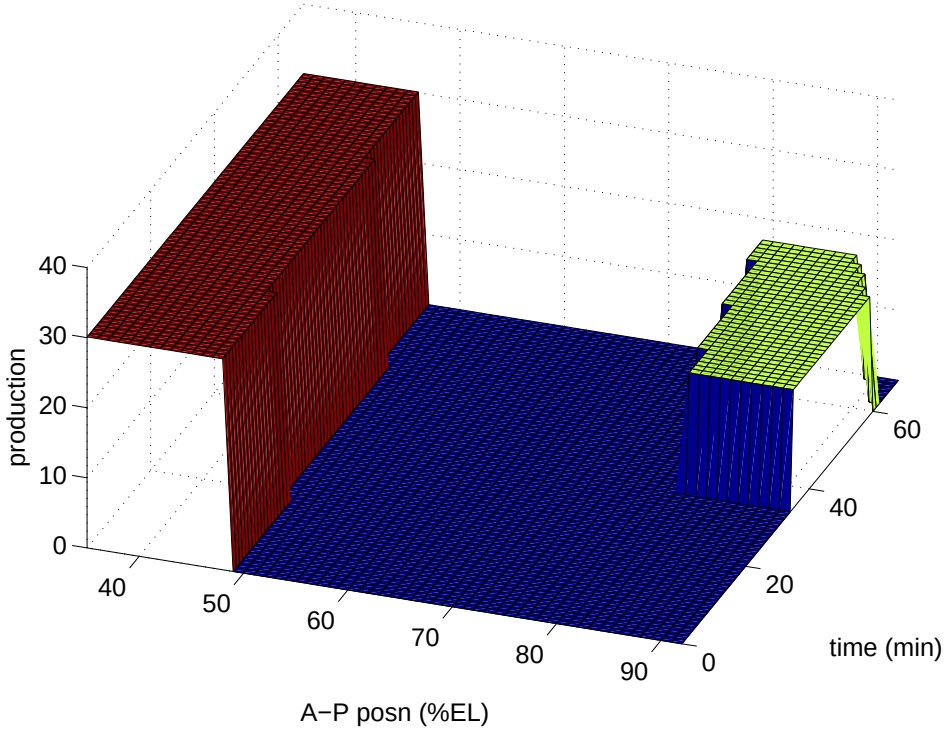


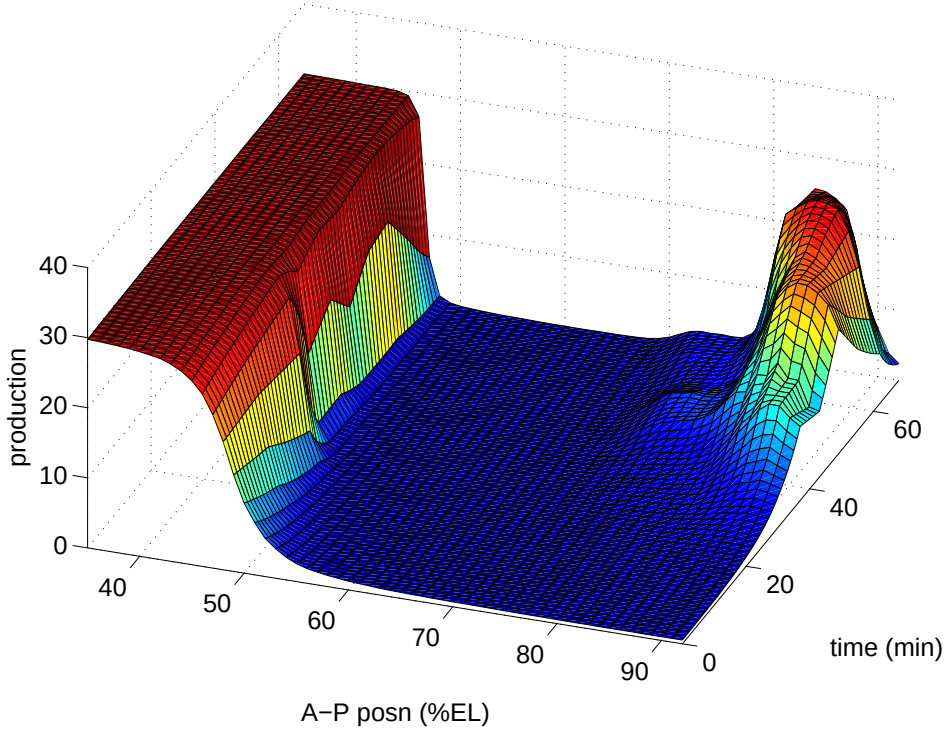


Step 2: Estimate R^a, T^{ab}, h^b based on $P^a(x,t)$

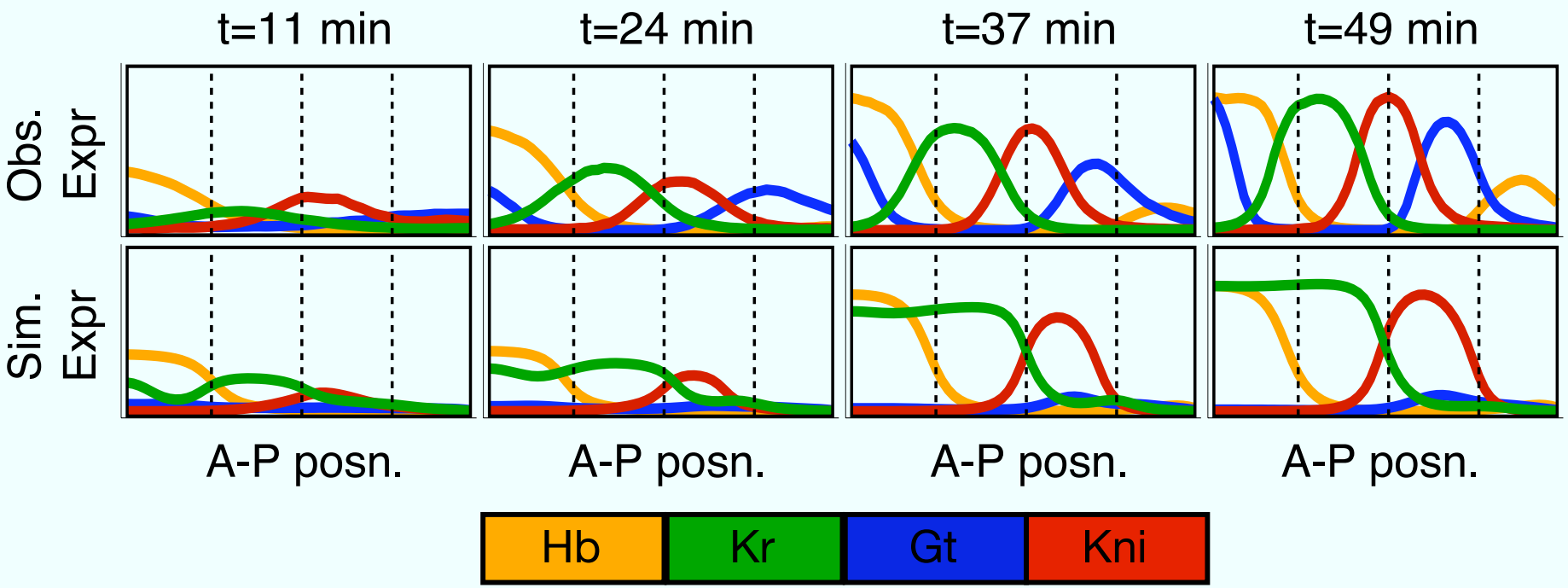
$$P^a(x,t) = R^a g\left(\sum_b T^{ab} v^b(x,t) + h^a\right)$$

- Repeated gradient descent to minimize sum squared error



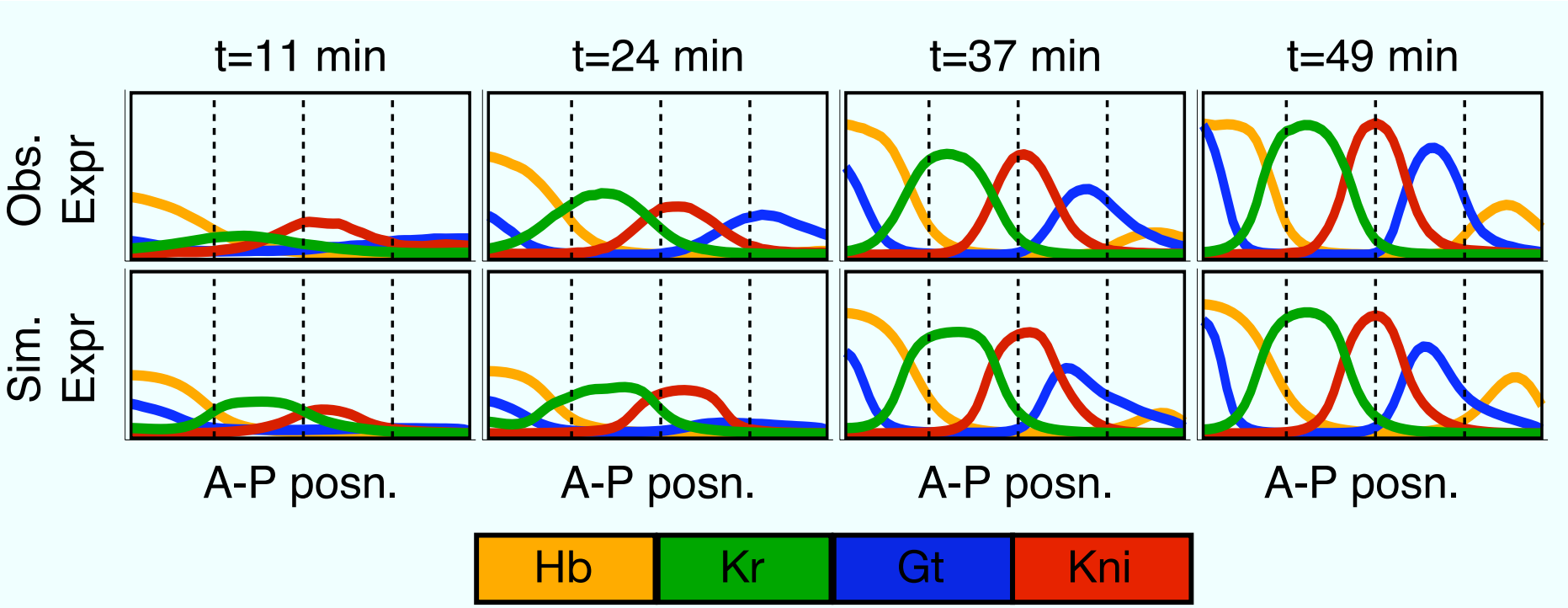


Simulating PDE with $R^a, T^{ab}, h^a, \gamma^a, D^a$ gives poor fit



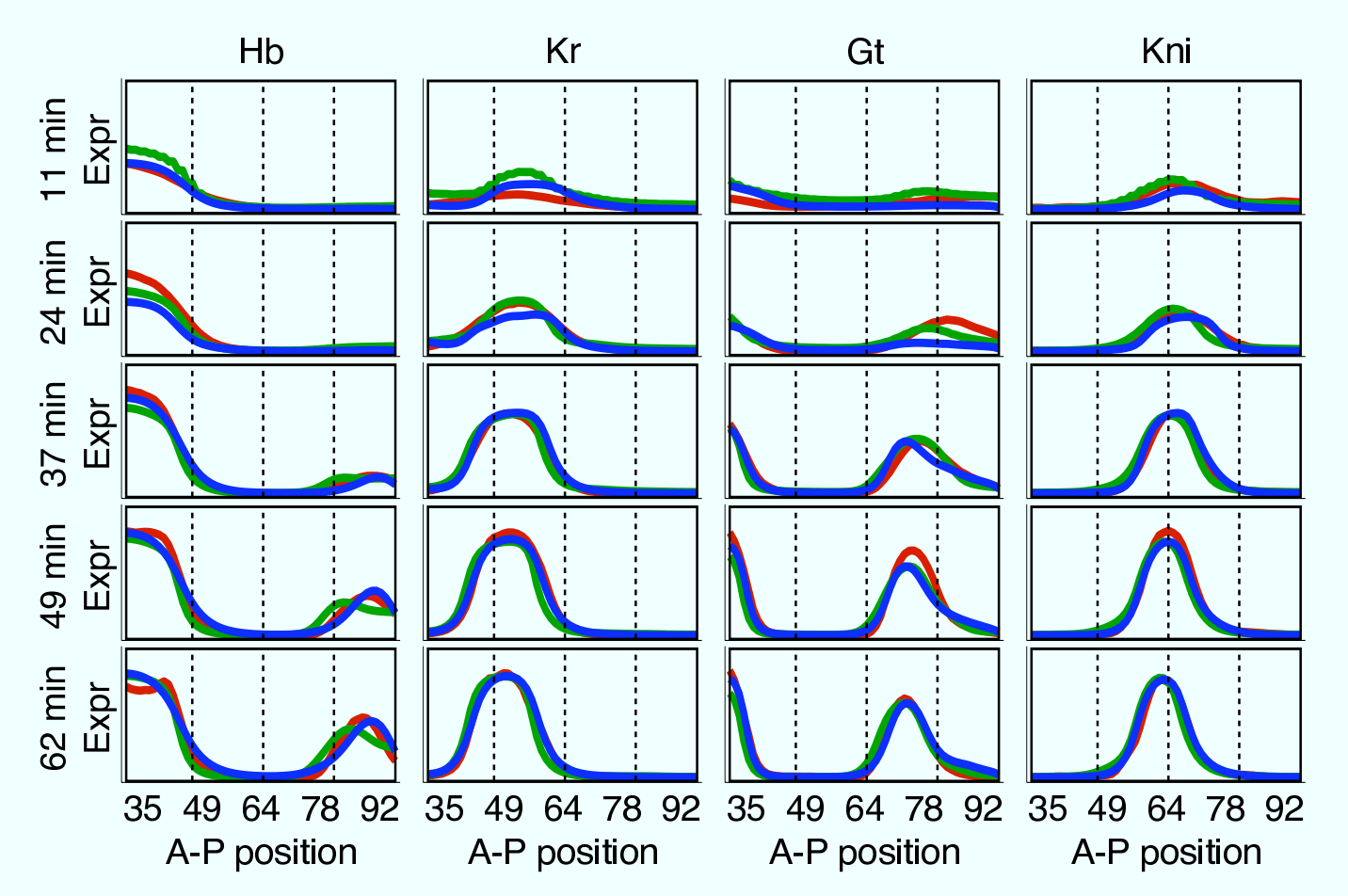
Step 3: Tune $R^a, T^{ab}, h^a, \gamma^a, D^a$ to get good fit

- Repeated stochastic local search



Results

Obtained similar quality to Jaeger et al. (2004a,b). . . in 36 hours!



Data (red); Jaeger et al. (green, RMS 12.08); Our fit (blue, RMS 12.29)

Results

Obtained similar architecture to Jaeger et al. (2004a,b)

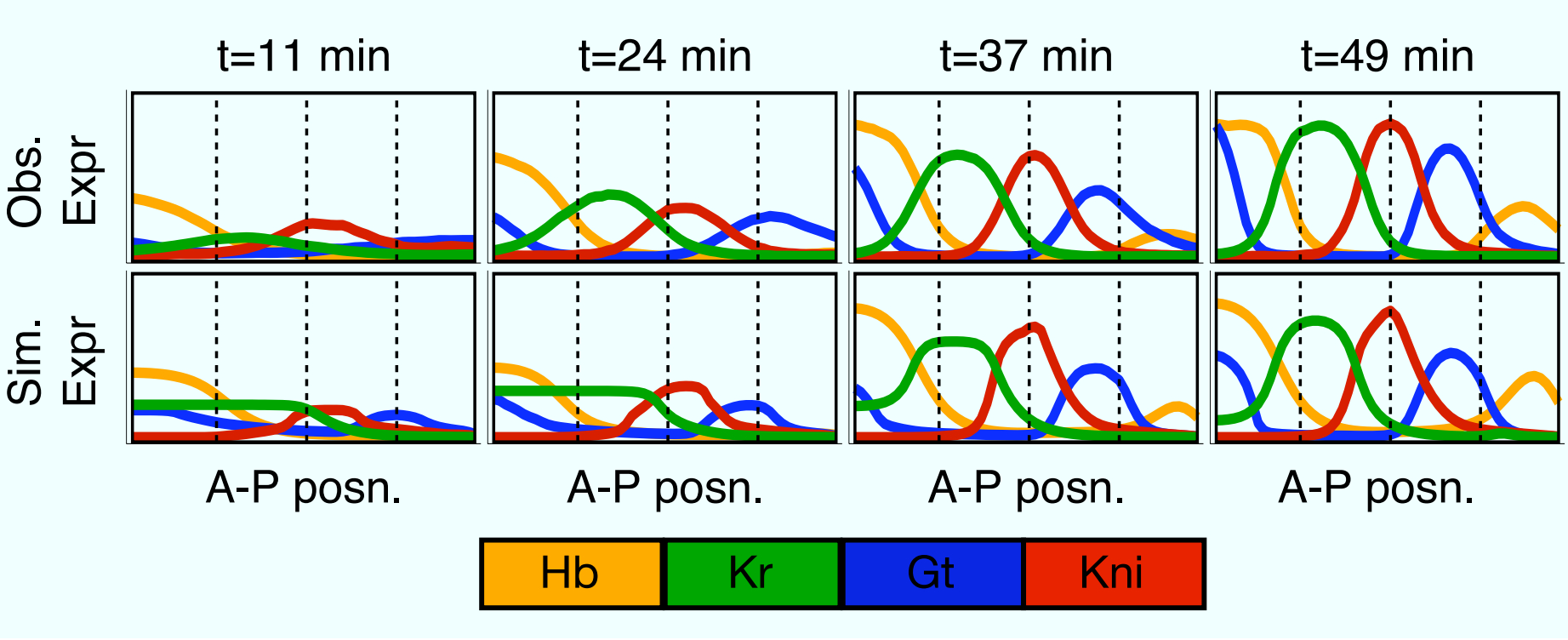
Network	Hb regulation Bcd Cad Hb Kr Gt Kni Tll	Kr regulation Bcd Cad Hb Kr Gt Kni Tll	Gt regulation Bcd Cad Hb Kr Gt Kni Tll	Kni regulation Bcd Cad Hb Kr Gt Kni Tll
R-P & J	$+\cdot+-\cdot\cdot\cdot$	$+\cdot\mp\cdot---$	$++\cdot-\cdot--$	$++-+-\cdot-$
Jaeger et al.	$+++0+-0$	$++-+---$	$++--+0-$	$++----+-$
Unc-GC	$+-+-+-+$	$++-+---$	$++--++-$	$++----+-$

Aside: One thing we seem to get right: post-Hb domain activated by Tll and sustained by Hb autoactivation.

Is that the only regulatory architecture that works?

- Next, we fit a model of the same form but limited to the RPJ regulatory relationships
- Regulatory weights T^{ab} corresponding to links not in the RPJ model are fixed at zero
- Regulatory weights T^{ab} corresponding to link in the RPJ model are constrained to have the appropriate sign
- A few exceptions:
  - We allowed Tll to activate Hb
  - There was an extra negative weight T^{Kr,Hb^2} multiplied by $(v^{Hb}(x,t))^2$, to allow Hb to have a dual regulatory effect on Kr

Model restricted to RPJ structure



RMS error 15.88

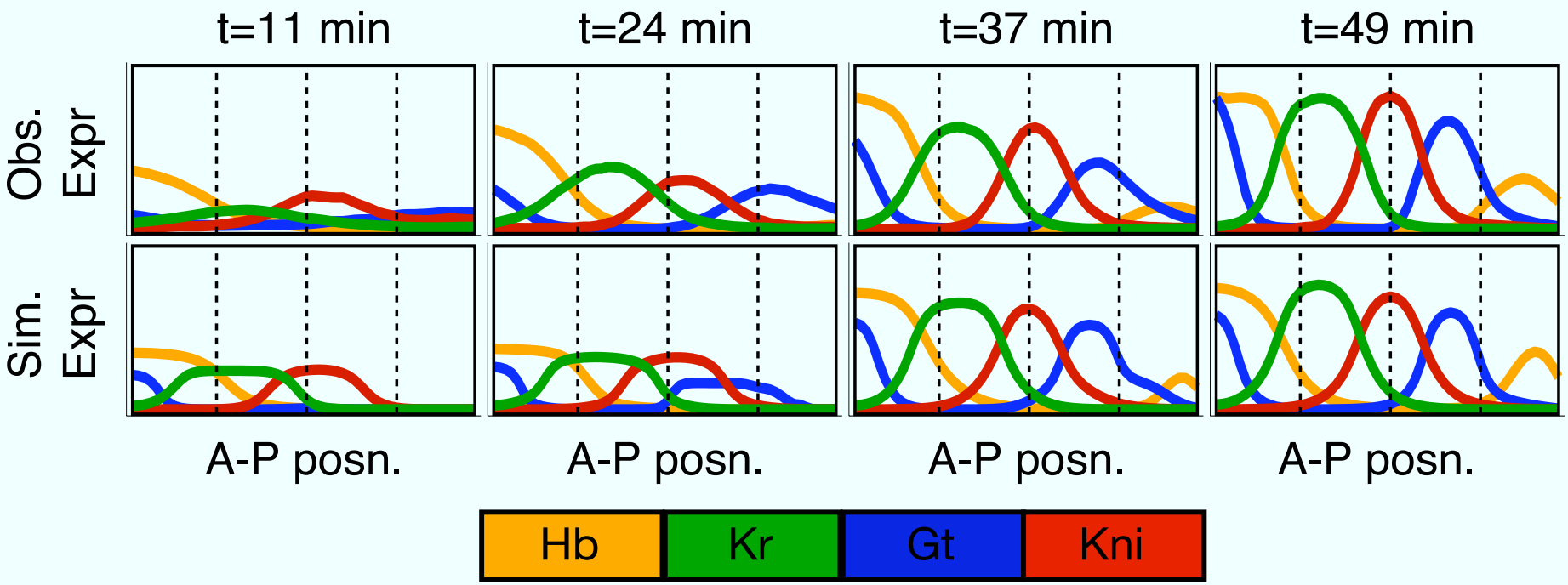
Does the mathematical form of the model matter?

- Next, we fit a piecewise-constant (“logical”) model for production
- We assumed production if at least one activator and no repressors exceed thresholds

$$P^{Hb} = \begin{cases} R^{Hb} & \text{if } (v^{Bcd} > 20 \text{ or } v^{Hb} > 90) \text{ and } v^{Kr} < 140 \\ & \text{and } v^{Kni} < 10 \\ 0 & \text{otherwise} \end{cases}$$

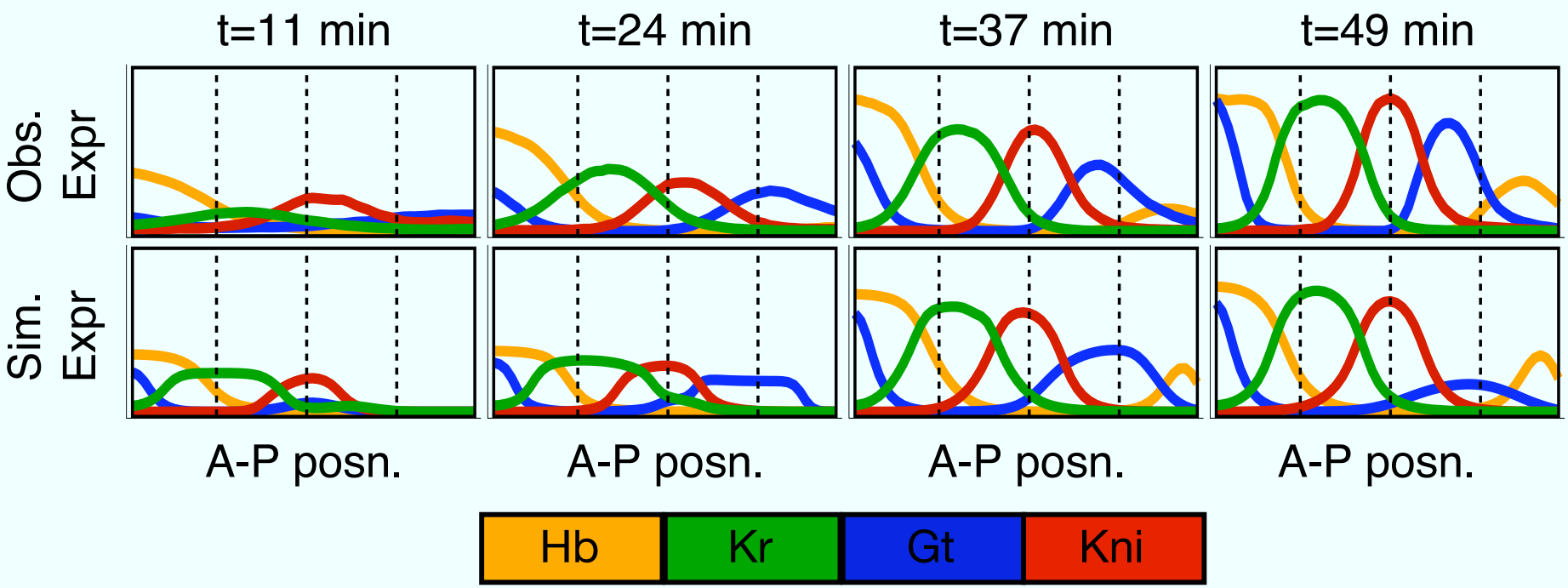
- Optimized thresholds, but not structure of network – we borrowed the structure of the first, unconstrained fit

Logical model with UncGC structure



RMS error 14.83

Logical model with RPJ structure



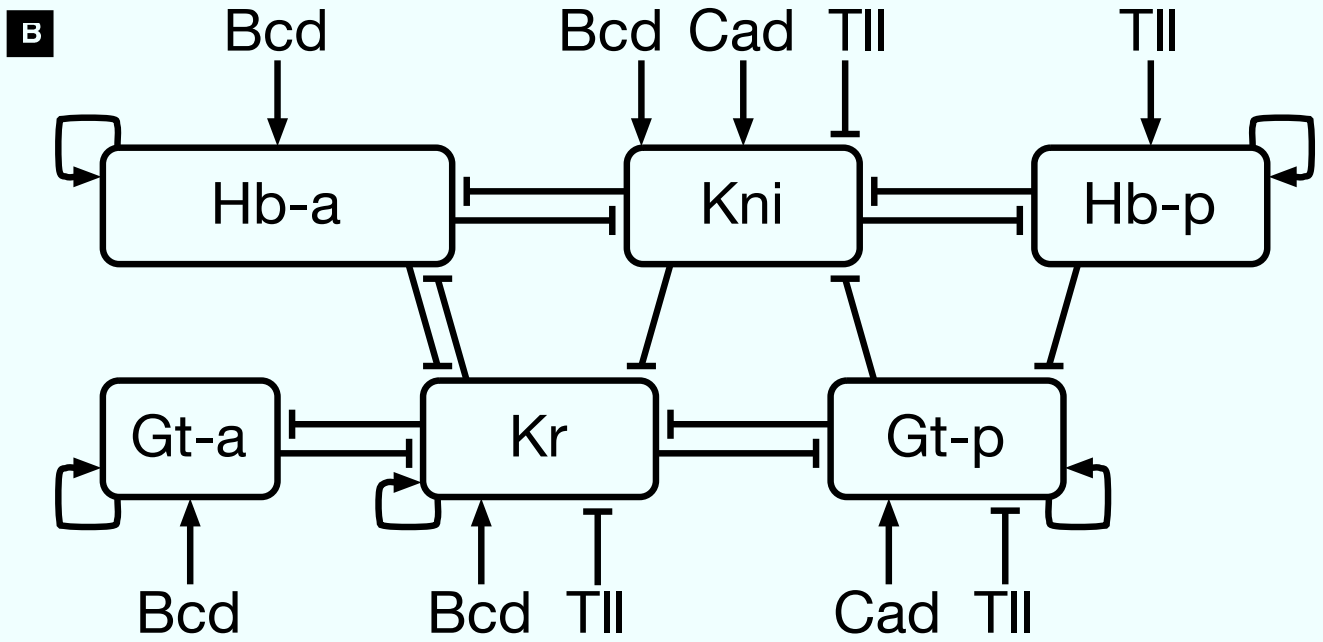
RMS error 21.91

“Consensus” network model

A

Network	Hb regulation							Kr regulation							Gt regulation							Kni regulation						
	Bcd	Cad	Hb	Kr	Gt	Kni	Tll	Bcd	Cad	Hb	Kr	Gt	Kni	Tll	Bcd	Cad	Hb	Kr	Gt	Kni	Tll	Bcd	Cad	Hb	Kr	Gt	Kni	Tll
Unc-GC	+	−	+	−	+	−	+	+	+	−	+	−	−	−	+	+	−	−	+	+	−	+	+	−	−	−	+	−
Unc-Logic	+	−	+	−	·	−	+	+	⊕	⊖	⊕	−	−	⊖	+	+	⊖	−	+	·	⊖	+	⊕	−	⊖	−	⊕	⊖
RPJ-GC	+	·	+	−	·	·	+	+	·	+−	·	−	−	−	+	+	·	−	·	⊖	−	⊕	+	−	⊕	−	·	−
RPJ-Logic	+	·	+	−	·	·	+	+	·	⊕−	·	−	−	−	+	+	·	−	·	⊖	−	+	⊕	−	⊕	⊖	·	⊖
Combined	+	·	+	−	·	−	+	+	·	−	+	−	−	−	+	+	−	−	+	·	−	+	+	−	·	−	·	−
R-P & J	+	·	+	−	·	·	·	+	·	+−	·	−	−	−	+	+	·	−	·	−	−	+	+	−	+	−	·	−
S & T	+	·	+	−	·	·	·	+	·	+−	·	−	−	·	+	+	−	−	·	·	·	+	+	−	·	−	·	·
Jaeger et al.	+	+	+	0	+	−	0	+	+	−	+	−	−	−	+	+	−	−	+	0	−	+	+	−	−	−	+	−

B



Conclusions of Part 2

- Quantitative model-fitting by simulated annealing (Reinitz)
  - Recovered *de novo* the regulatory architecture of the gap gene system
  - Revealed previously undetected & unexplained shifts in the expression domains of Kr, Kni and Gt
- Faster model fitting based on a hybrid of the regression approach and direct optimization (Perkins)
  - Essentially recovered Reinitz's results, but hundreds of times faster
  - May have explained formation of post-Hb correctly
  - There is not a unique regulatory architecture capable of accounting for the data
  - Alternative mathematical forms for the model can explain the data
- Future work includes relating the dynamical models to more biologically meaningful parameters related to promoter sequence

That’s all folks!