Systems Biology of Pattern Formation, Canalization and Transcription in the *Drosophila* Blastoderm

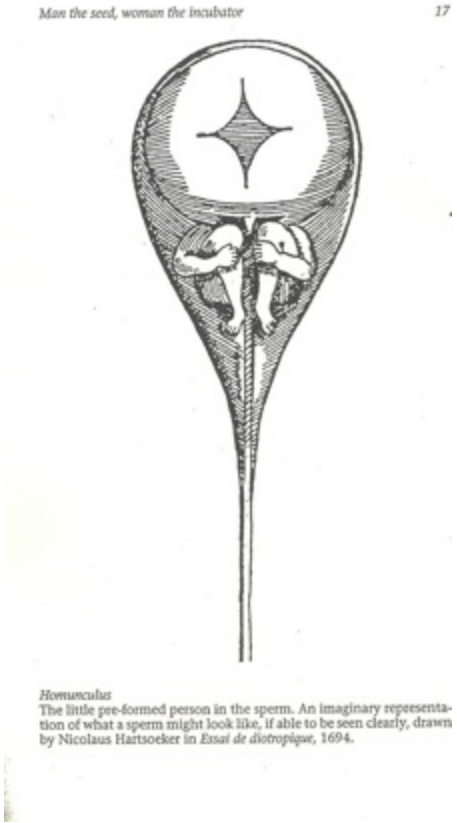
John Reinitz

STAT Applied Math Retreat

Gleacher Center

Motivation: Understanding how Biological Form is Created *de novo*?

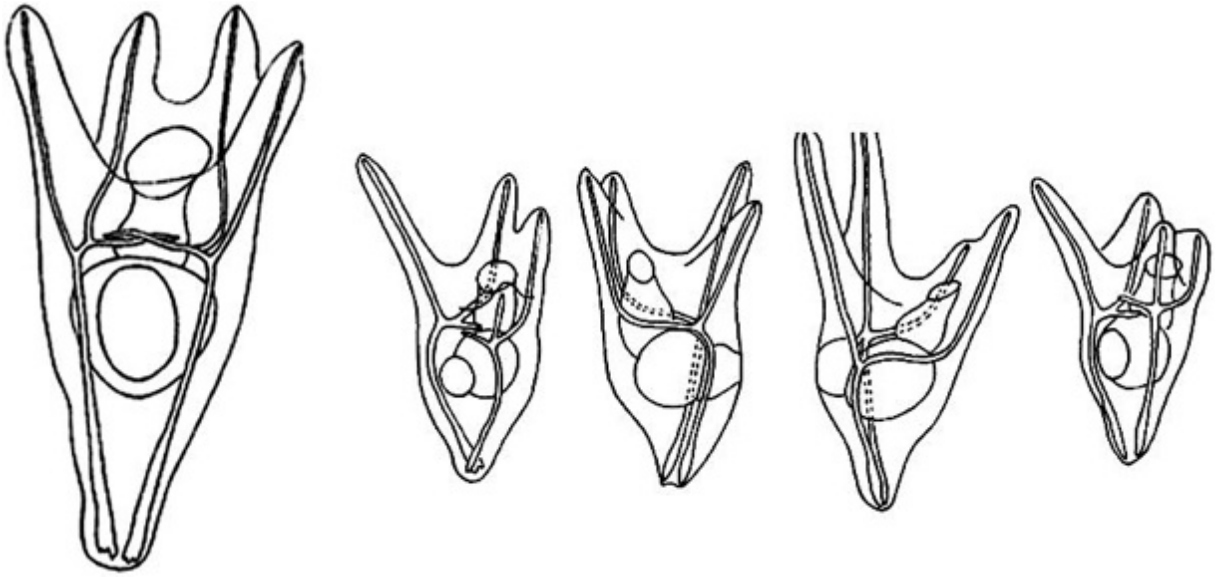
Hippocrates and Aristotle believed that form was present in miniature as a homuculus (later speculated to be in head of sperm by Hartsoeker, 1694): i.e. form cannot be created *de novo*; instead it is preformed.



Homunculus
The little pre-formed person in the sperm. An imaginary representation of what a sperm might look like, if able to be seen clearly, drawn by Nicolaus Hartsoeker in *Essai de dioptrique*, 1694.

Regulative Development

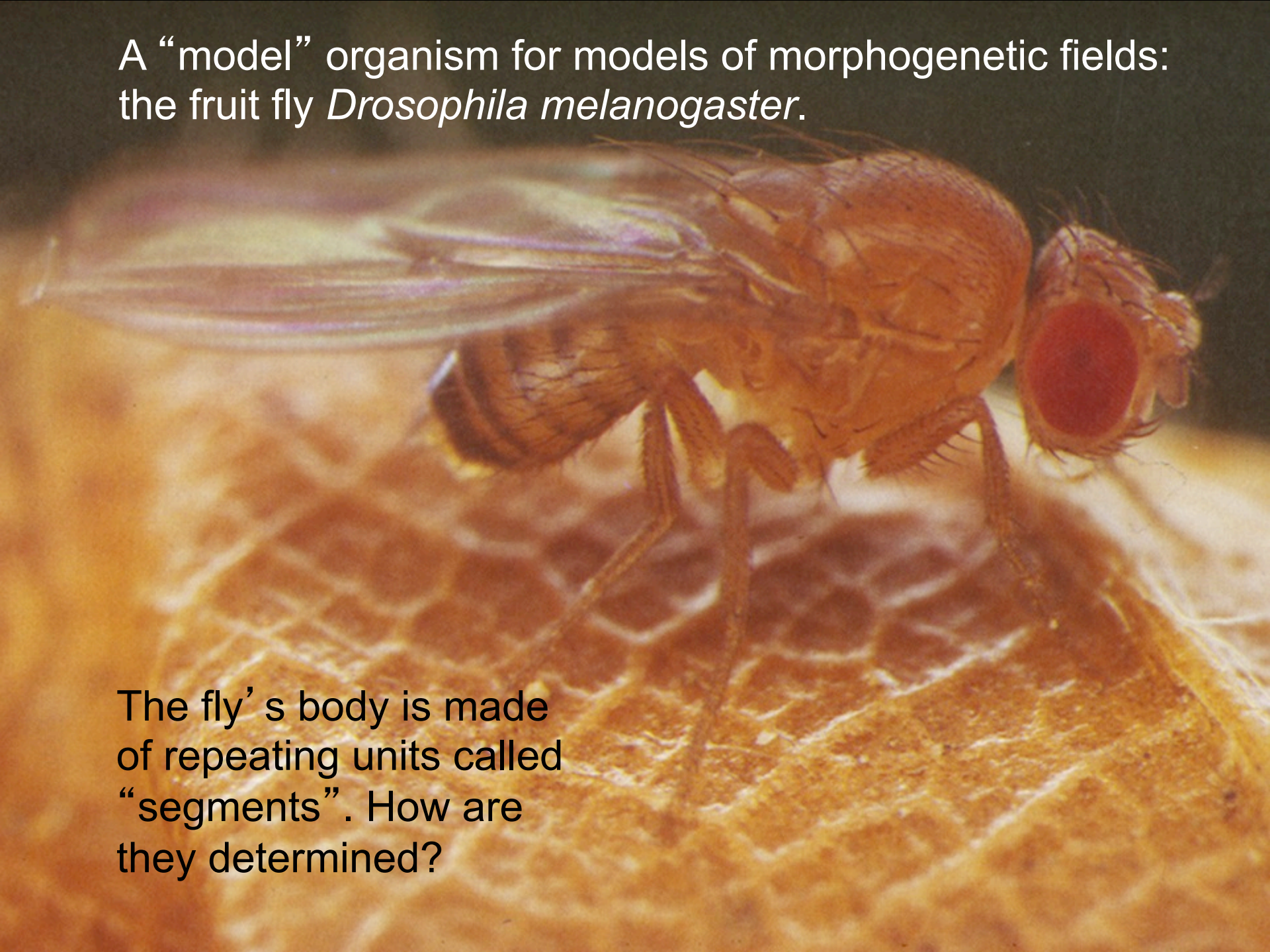
Driesch (1891) disproved preformationalism by showing that early sea urchin embryos dissociated into individual cells develop into whole sea urchins.



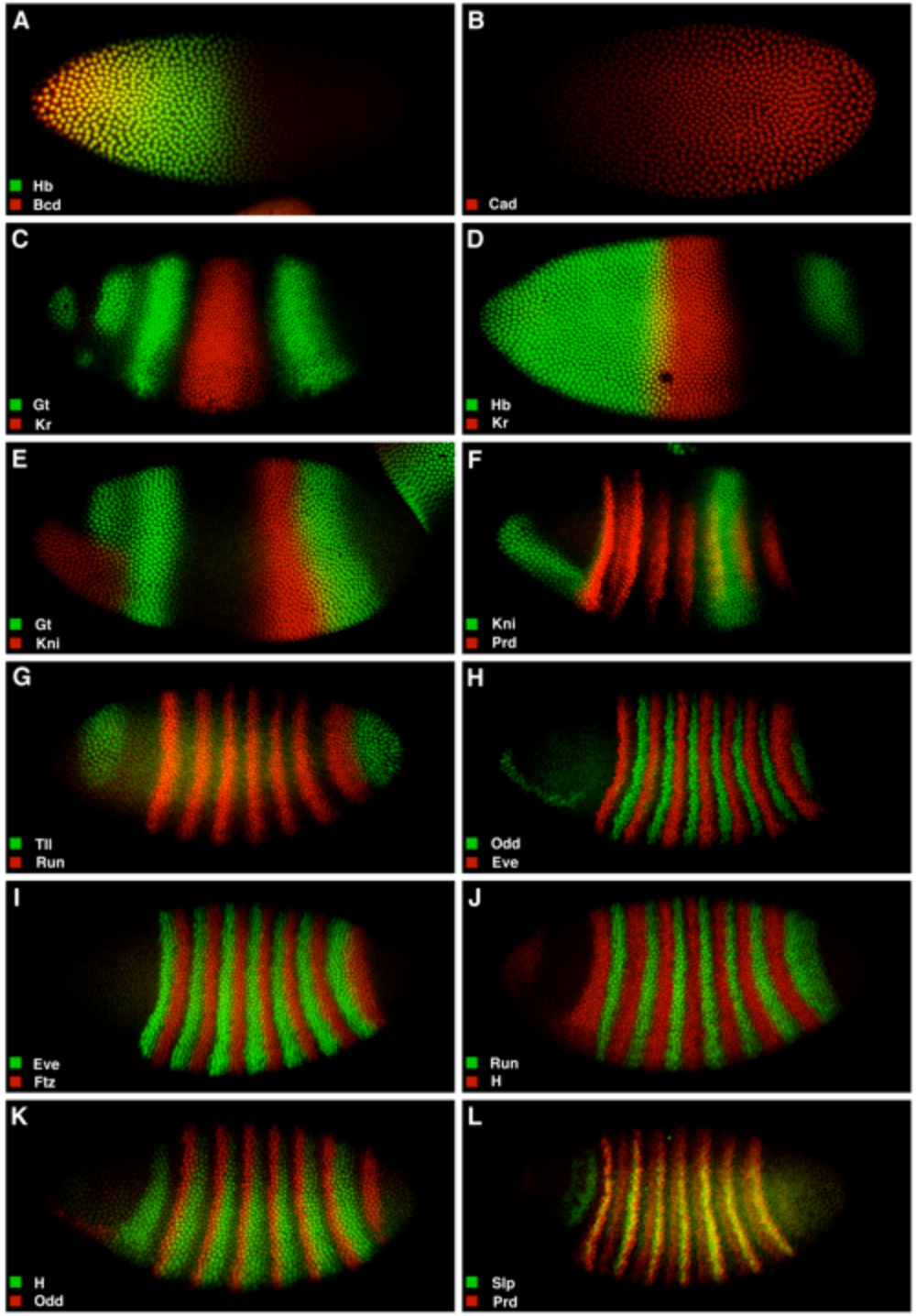
Hörstadius and Wolsky, 1936, W. Roux. Arch. *Entw. Mech. Org.* **135**:69–113 (1936), via DeRoberis *Science* **126**:925-941 (2009)

How to explain? Driesch gave up, but now these problems can be approached..

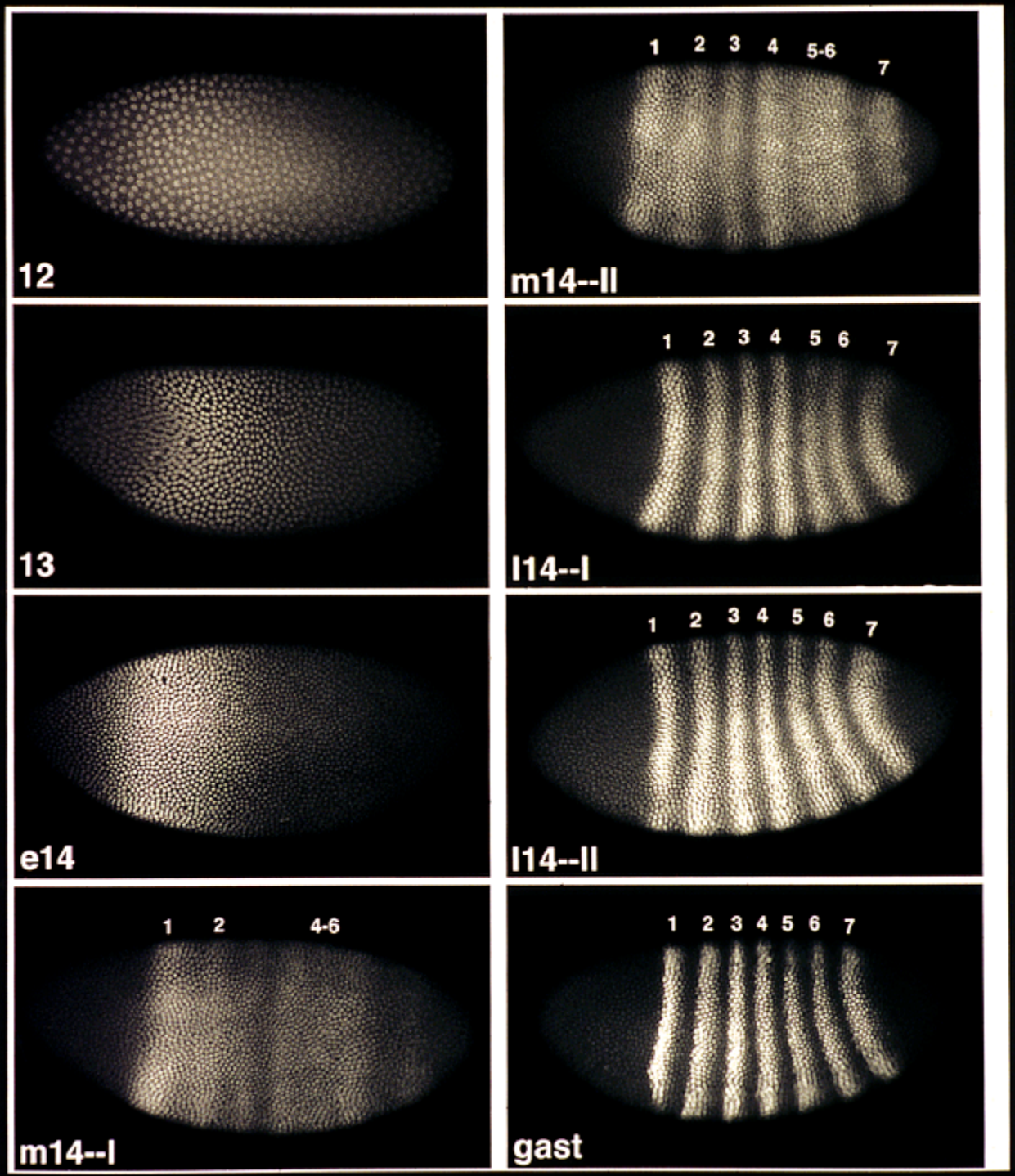
A “model” organism for models of morphogenetic fields: the fruit fly *Drosophila melanogaster*.



The fly’s body is made of repeating units called “segments”. How are they determined?



14 segmentation genes are expressed in the blastoderm. Each has a distinct pattern of expression.



In addition, each expression pattern changes over time.

Blastoderm Systems Biology: Three Central Problems

1. Pattern Formation. How are fates determined in the segment determination system of *Drosophila*? We use a differential equation model in conjunction with quantitative observations of gene expression.

2. Canalization. How are errors in development corrected?

3. Transcriptional Control. What are the fundamental rules that control how transcription of key developmental genes are controlled by binding sites, groups of which form modular enhancers? We use a quantitative model based on DNA sequence and observed expression. Applications include synthetic enhancers, driving (we hope!) arbitrary expression patterns. Also want to apply the model to the study of polymorphisms: can we make GWAS unnecessary?

Genes: Key Properties for Pattern Formation & Error Correction

Genes are very complex things, and are treated in different ways in different contexts. For understanding development, the key properties of genes are the following:

1. Each gene is able to synthesize a protein (by first making RNA).
2. A gene may be synthesizing a protein at a given time ("turned on") or it may be inactive ("turned off").
3. Some genes make proteins whose biological function is to turn other genes on and off.
4. In a multicellular organism (like you, me, and a fly), all cells have the same genetic material, but different genes are turned on in different types of cells (skin, muscle, etc).

Synthesis

$$\frac{dv_i^a}{dt} = R_a g_a \left(\sum_{b=1}^{N} T^{ab} v_i^b + m^a v_i^{\mathrm{Bcd}} + h_a \right)$$

Transport

$$+ D^a(n) \left[\left(v_{i-1}^a - v_i^a \right) + \left(v_{i+1}^a - v_i^a \right) \right]$$

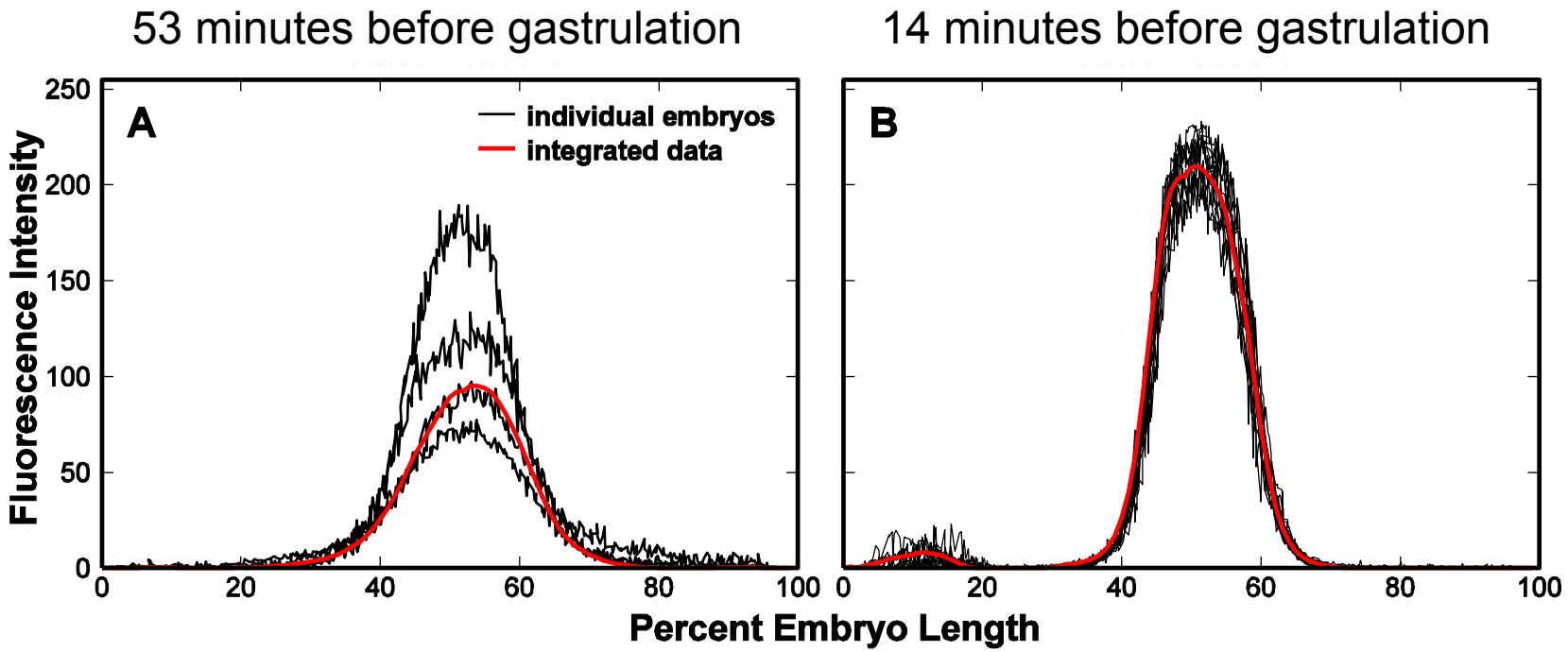
Decay

$$- \lambda_a v_i^a$$

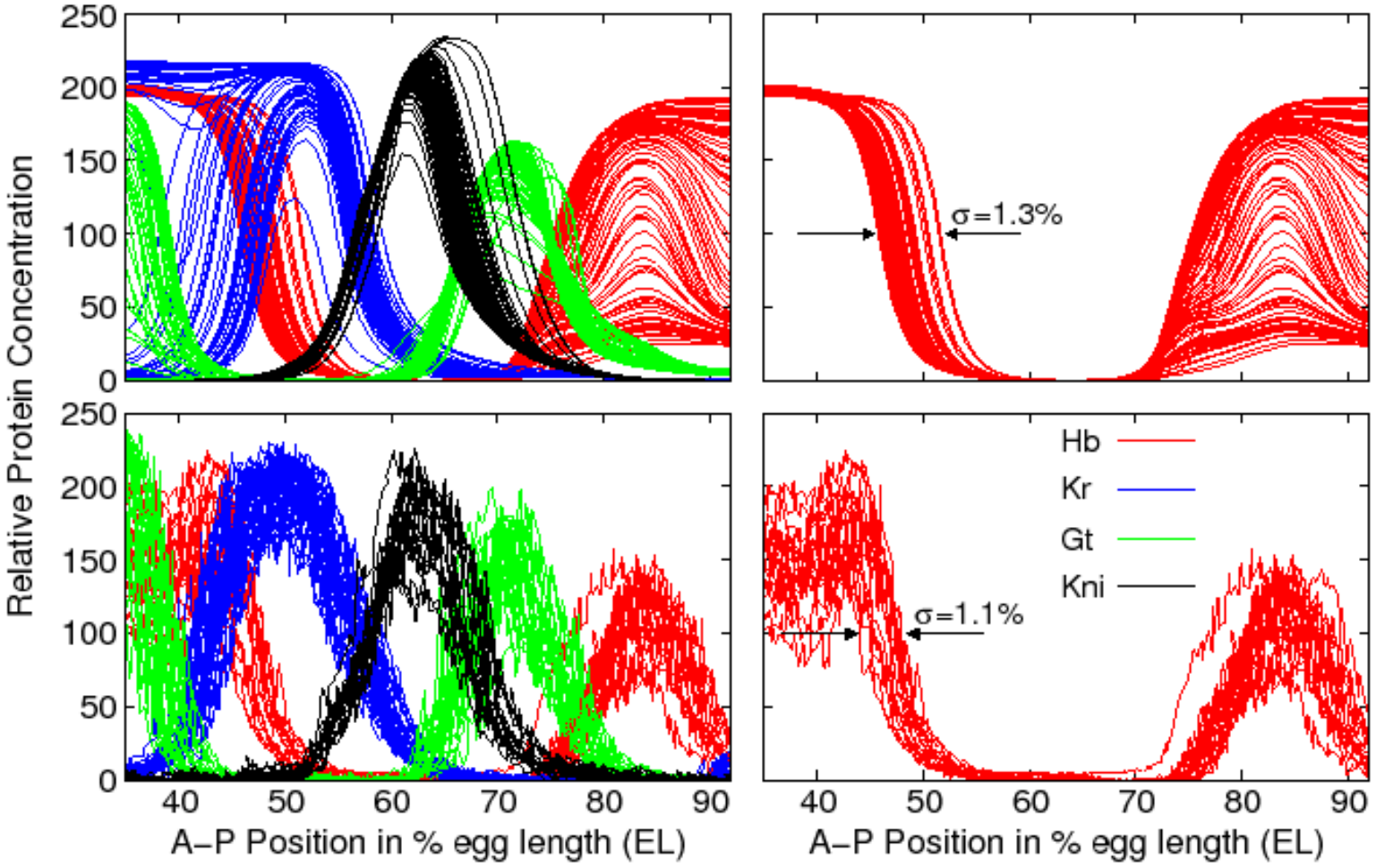
General Result: Patterns are Variable at Early Times, Uniform by Gastrulation

This is the molecular implementation of canalization.

(Shown: *Kruppel* expression in 1D strip at central10% of dorsal-ventral coordinates)

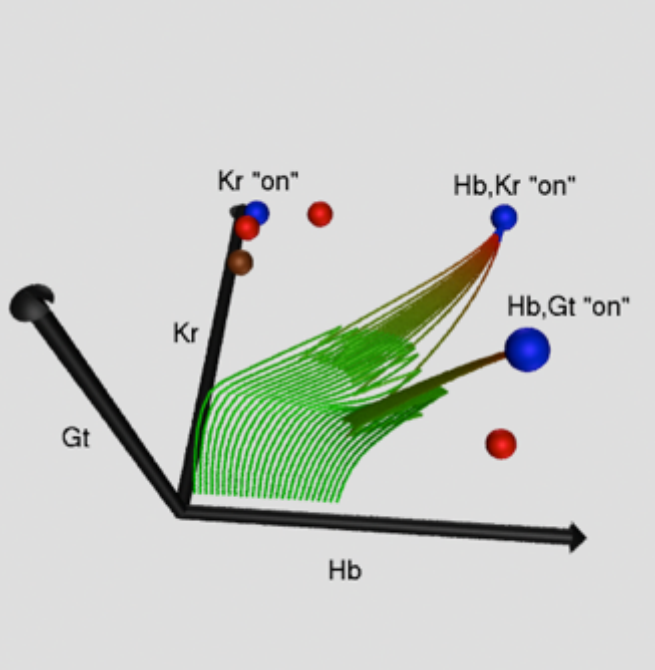


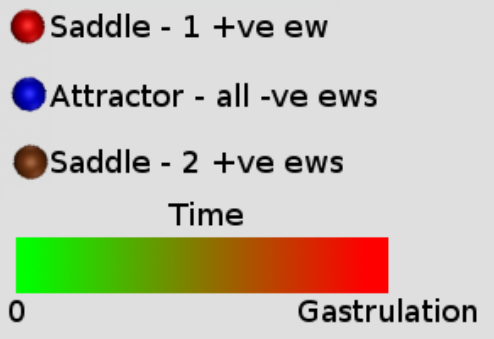
Model correctly predicts reduction of variance

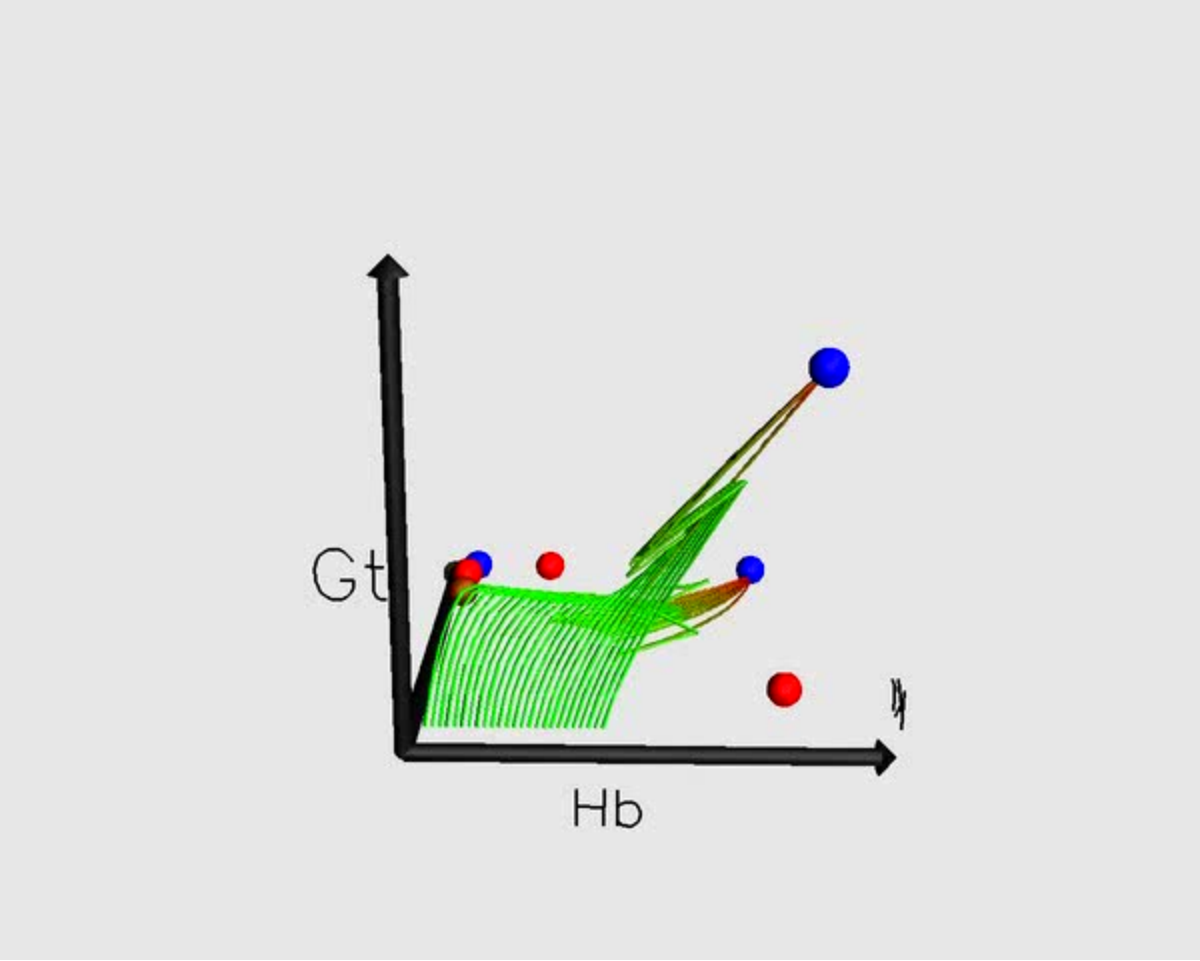


2. Canalization: Dynamical Analysis

Dynamical structure in the anterior: canalization by point attractors.







What is the Genetic Code for Regulation?

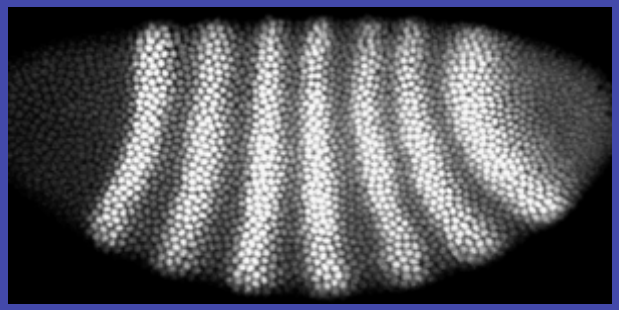
Recall from previous slide that:

1. A gene may be synthesizing a protein at a given time ("turned on") or it may be inactive ("turned off").
2. Some genes make proteins whose biological function is to turn other genes on and off.

The big question:

How is regulatory logic implemented in noncoding DNA sequence?

The *even-skipped* genomic region and key enhancers



Sequence level transcription model – equations

Janssens et al., Nature Genetics, **38**:1159-65 (2006); Kim et al., PLoS Geneics **in press**

DNA sequences
PWMs
Concentrations of Trans-acting factors

$$S_i^a = \sum_{k=1}^{l} \ln\left(\frac{w_b^k}{p_b}\right) \quad (1) \rightarrow K_{i[m_i,n_i;a]} = K_{\max}^a \exp\left(\frac{S_i^a - S_{\max}^a}{\lambda}\right) \quad (2)$$

Fractional occupancies of DNA binding factors

$$f_{i[m_i,n_i;a]} = \frac{K_{i[m_i,n_i;a]} v^a}{1 + K_{i[m_i,n_i;a]} v^a} \quad (3)$$

$$f_{i[m_i,n_i;a]} = \frac{K_{i[m_i,n_i;a]} v^a}{1 + K_{i[m_i,n_i;a]} v^a + K_{j[m_j,n_j;b]} v^b} \quad (4)$$

$$f_{i[m_i,n_i;a]} = \frac{K_{i[m_i,n_i;a]} v^a (1 + K_{j[m_j,n_j;b]} v^b K_{ab}^{coop})}{1 + K_{i[m_i,n_i;a]} v^a (1 + K_{j[m_j,n_j;b]} v^b K_{ab}^{coop}) + K_{j[m_j,n_j;b]} v^b} \quad (5)$$

Coactivation & Quenching & Direct Repression

$$f^Q_{i[m_i,n_i;a]} = f^T_{i[m_i,n_i;a]} \prod_k (1 - c(d_{ik}) E_b^{CA} f_{k[m_k,n_k;b]})$$
$$f^A_{i[m_i,n_i;a]} = f^T_{i[m_i,n_i;a]} - f^Q_{i[m_i,n_i;a]} \quad (6)$$

$$F^A_{i[m_i,n_i;a]} = f^A_{i[m_i,n_i;a]} \prod_k (1 - q(d_{ik}) E_b^Q f^Q_{k[m_k,n_k;b]}) \quad (7)$$

$$F^{AF} = f^{AF} \prod_k (1 - q(d_k) E_b^D f^Q_{k[m_k,n_k;b]}) \quad (8)$$

Integration of activation inputs

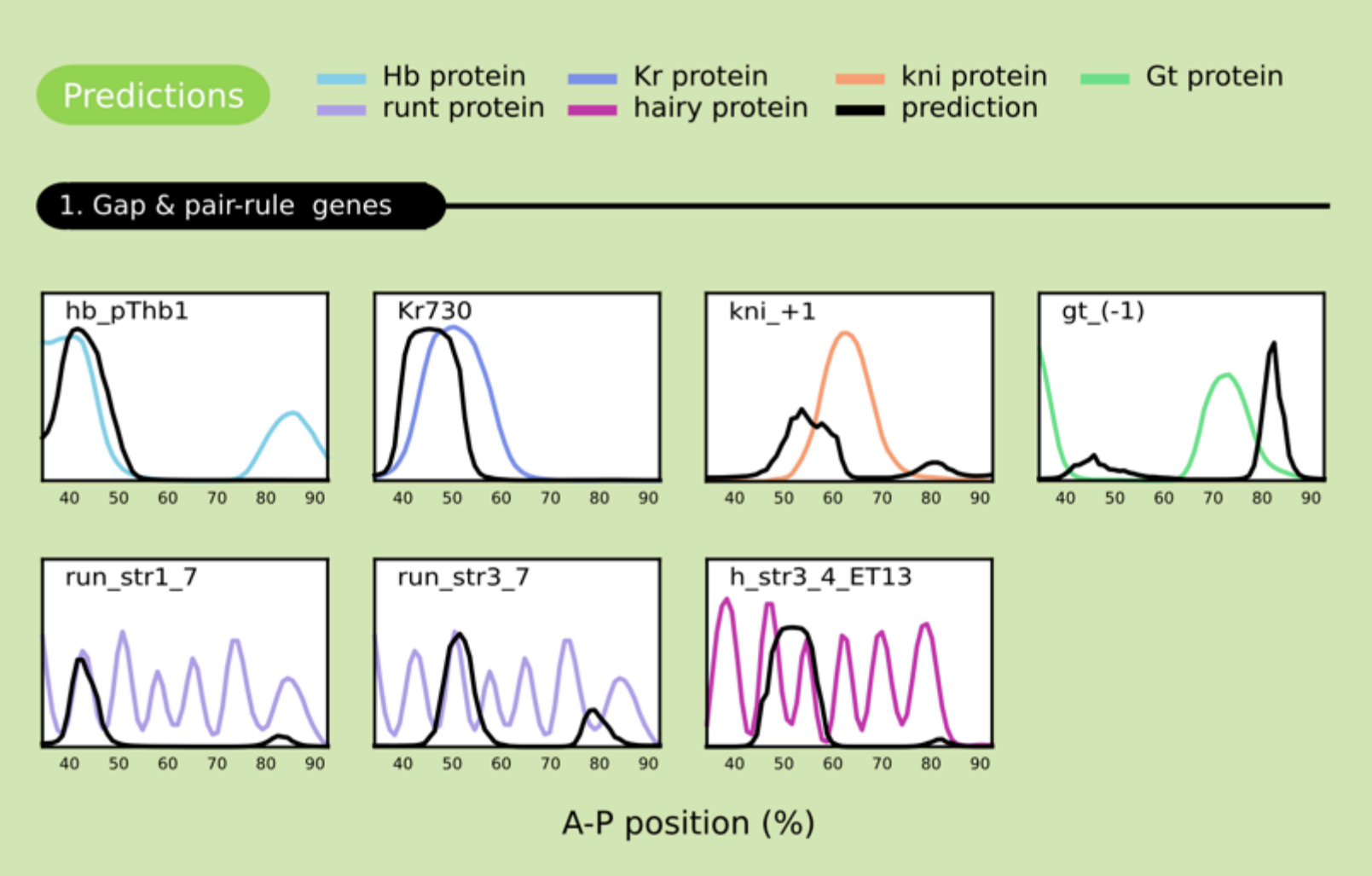
$$N = \sum_a E_a^A \sum_i F^A_{i[m_i,n_i;a]} \quad (9)$$

$$M = F^{AF} N \quad (10)$$

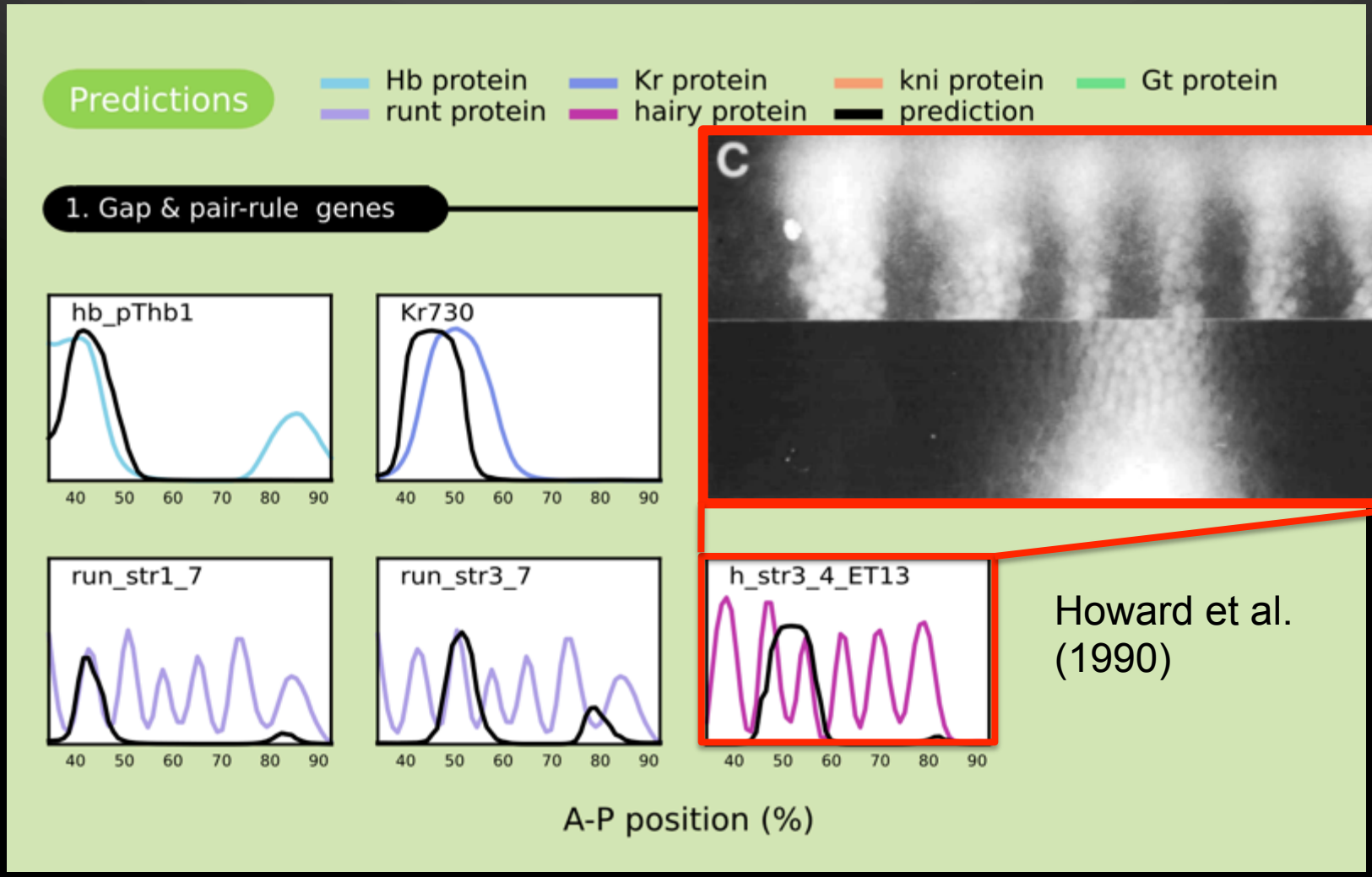
Transcription rates

$$\frac{d[mRNA]}{dt} \cong \frac{R_{\max} \exp(QM - \theta)}{1 + \exp(QM - \theta)} \quad (11)$$

Gap, pair-rule gene, other species prediction

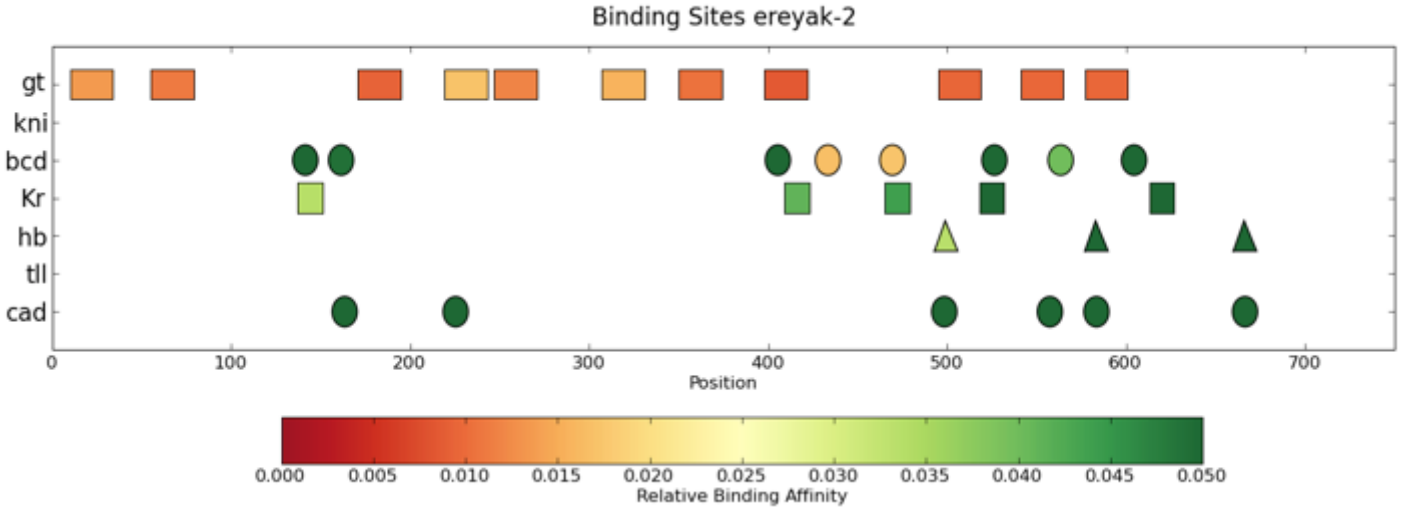


Gap, pair-rule gene, other species prediction

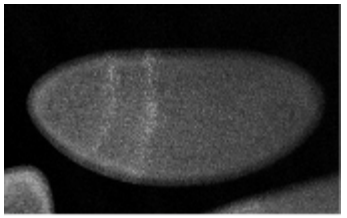


Howard et al. (1990)

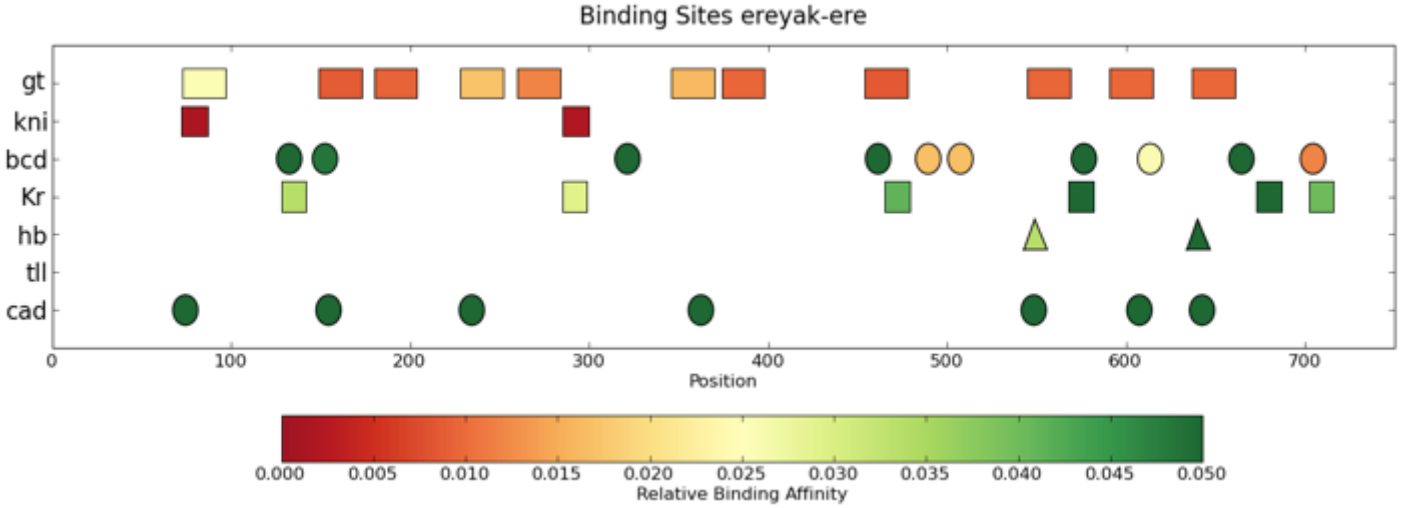
First Synthetic Enhancers



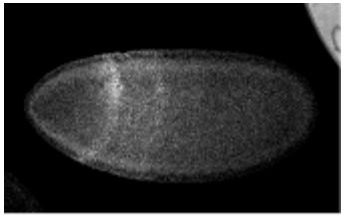
ereyak-2



positive control (98% homology to mel)



ereyak-ere



midway point between ereyak and ere (85% homology to mel)

nonhomologous synthetic constucts without competitive binding sites now being synthesized.

A course, next offered SpQ 2013
"Gene Regulation"
STAT/MGCB/ECEV 35400

- •math methods: Nonlinear ODE's, statistical physics of gene reg.
- •synthetic gene circuits: engineered circuits in coli., Drosophila
- •key problems for phage lambda and *Drosophila* solved with mathematical models.
- •how to design new theories for different types of problems.

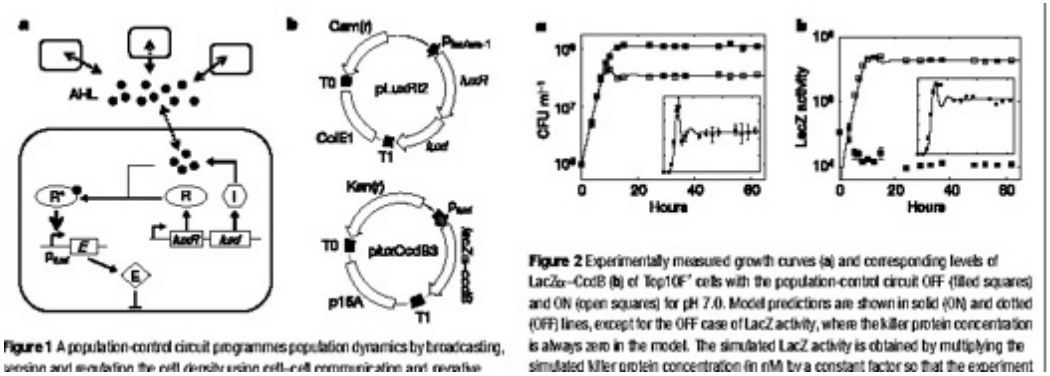


Figure 1 A population-control circuit programmes population dynamics by broadcasting,

Figure 2 Experimentally measured growth curves (a) and corresponding levels of LacZα–CcdB (b) of Top10F′ cells with the population-control circuit OFF (filled squares) and ON (open squares) for pH 7.0. Model predictions are shown in solid (ON) and dotted (OFF) lines, except for the OFF case of LacZ activity, where the killer protein concentration is always zero in the model. The simulated LacZ activity is obtained by multiplying the simulated killer protein concentration in nM by a constant factor so that the experiment

You et al., Nature **428**:868 (2005)