System biology of development

Maria Samsonova

(St.Petersburg Polytechic University, SPb, Russia)

How does the fertilized egg give rise to the adult body and how does adult body produce yet another body?

Three general questions:

- The question of differentiation: Since each cell of the body contains the same set of genes, how can this same set of genetic instructions produce different types of cells?
- The question of morphogenesis: The organs are spatially arranged in a particular way. How can the cells form such ordered structures ?
- The question of growth: How is cell division so tightly regulated?

Molecular and Spatial Differentiation

Molecular differentiation places the emphasis on the changes occuring within the cell with time and is mainly concerned with the control of the synthesis of specific macromolecules which are characteristic of the cell type.

Spatial differentiation precedes formation of tissue or organ. It is the process by which the individual cells within a population are specified to undergo a particular molecular differentiation, which results in a chararcteristic spatial pattern.

Molecular Differentiation

Cell commitment: the overt changes in cellular biochemistry and function are preceded by a process resulting in commitement of a cell to a certain fate.

The process of **commitment** can be divided into two stages:

Specification. The fate of a cell is said to be specified when it is capable of differentiating autonomously when placed in a neutral environment, such as petri dish or a tube.

Determination. A cell or tissue is said to be determined when it is capable of differentiating autonomously even when placed into another region of the embryo.

More Definitions

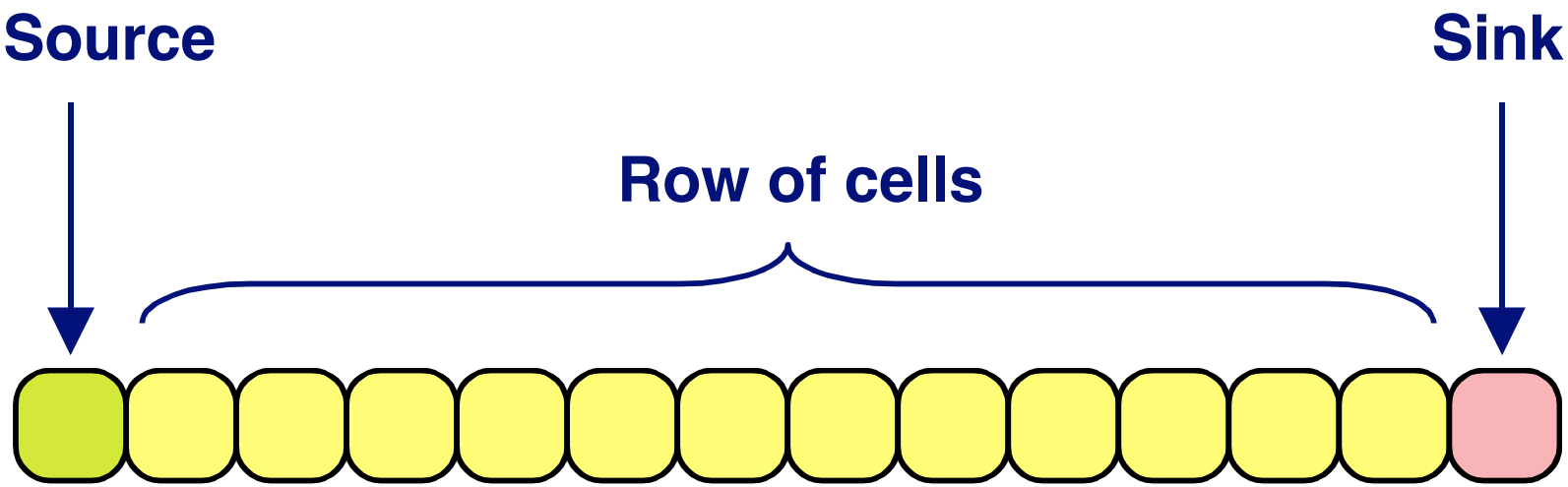
Positional Information

Mechanism whereby cells in a developing system have their position specified with respect to one or more points in the system.

Developmental Field

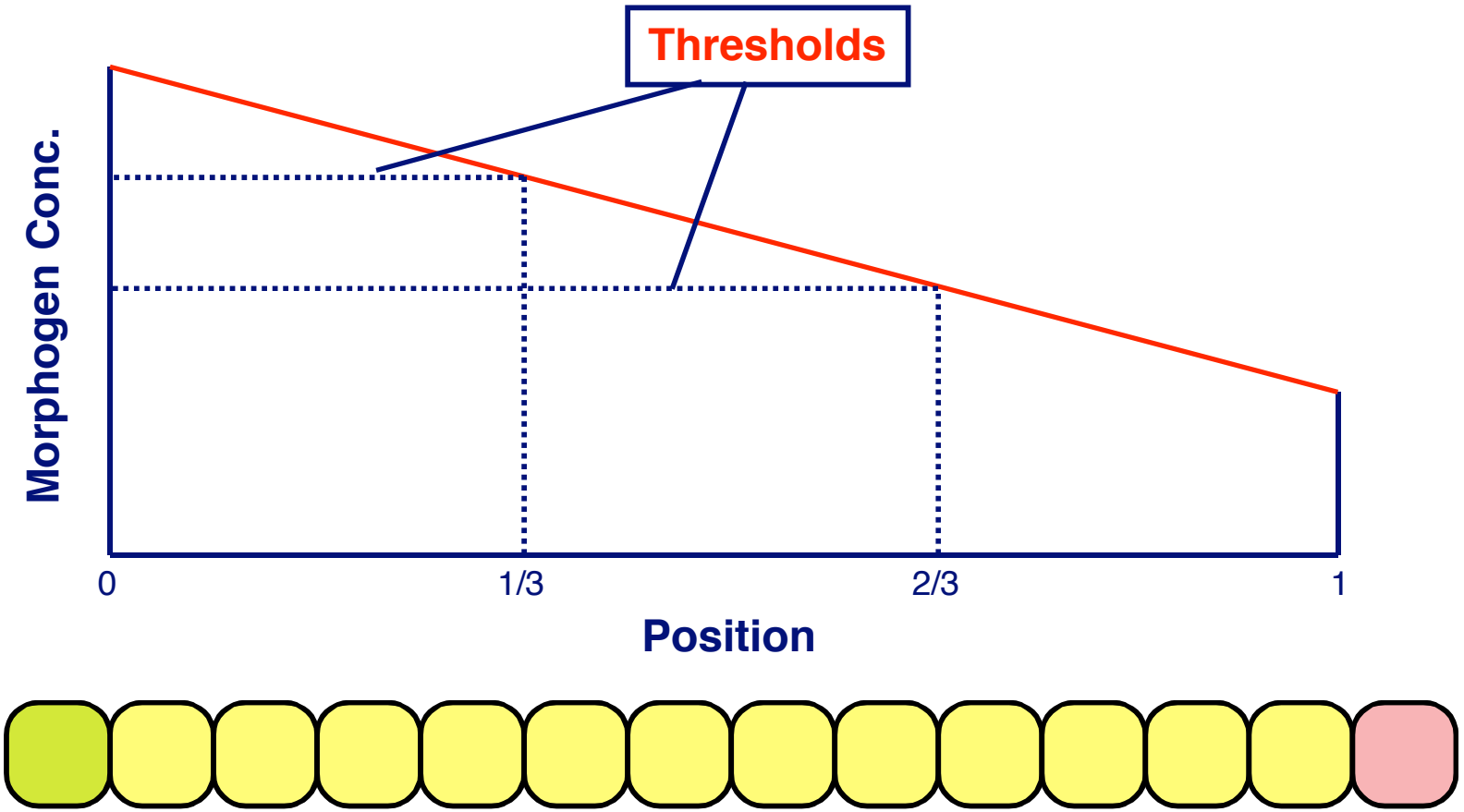
When cells have their positional information specified with respect to the same set of points, this constitutes a field.

The French Flag Problem



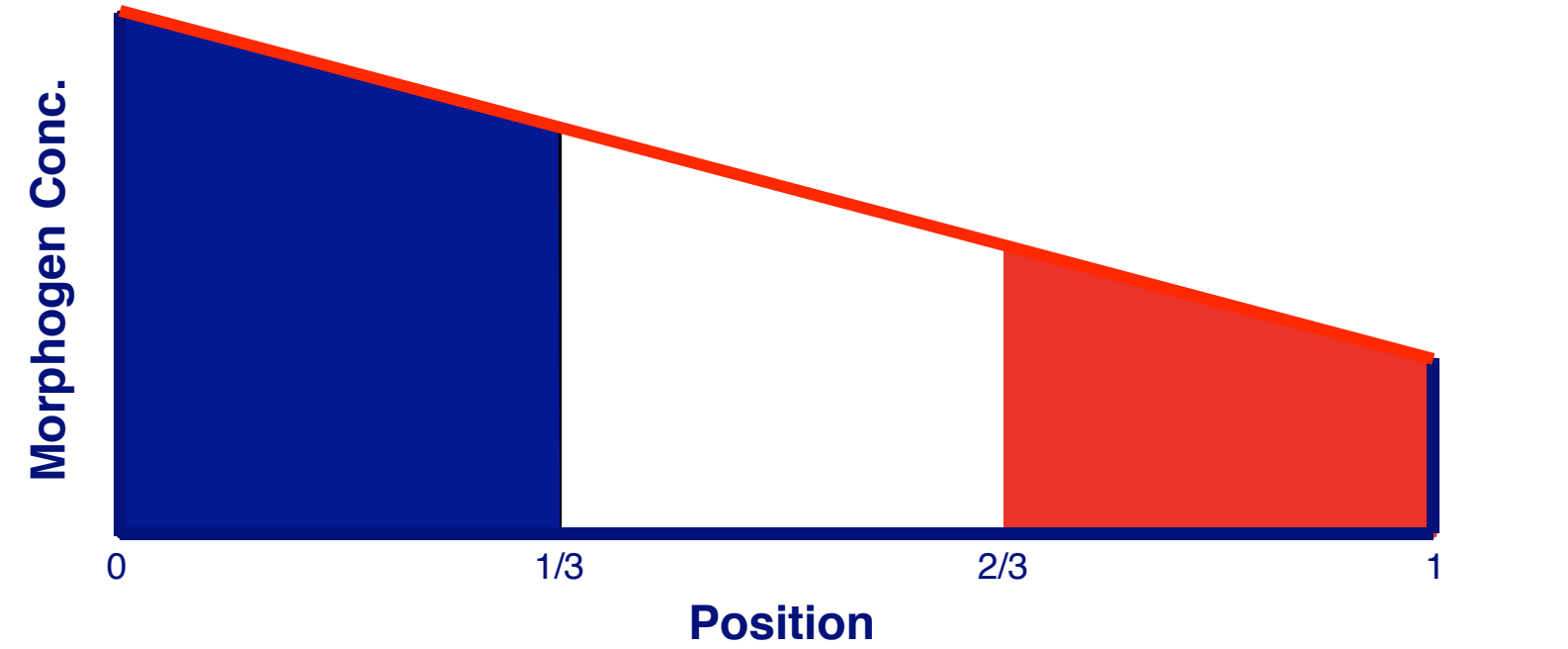
slightly modified from: Wolpert (1968). In: Towards a Theoretical Biology (Waddington ed.)

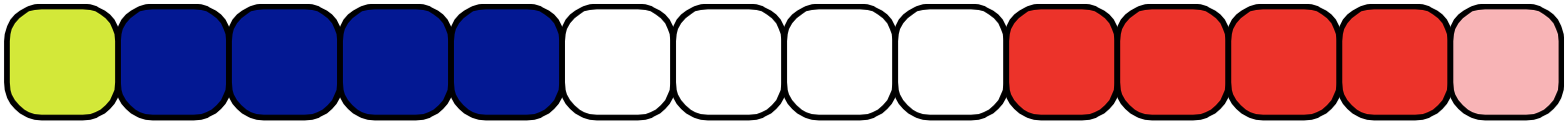
The French Flag Problem



slightly modified from: Wolpert (1968). In: Towards a Theoretical Biology (Waddington ed.)

The French Flag Problem





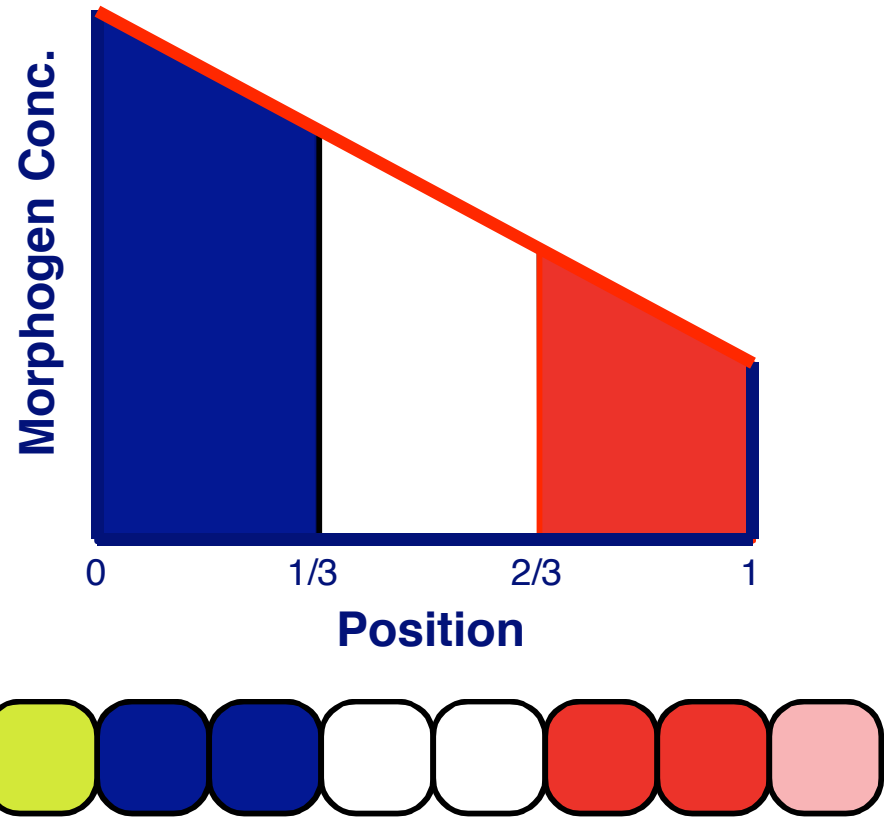
slightly modified from: Wolpert (1968). In: Towards a Theoretical Biology (Waddington ed.)

Positional Information

- largely determines (with respect to cell genome and developmental history) nature of molecular differentiation a cell will undergo;
- polarity is defined with respect to reference points; it is direction, in which positional information is specified or measured;
- can be chemical substance, potential or activity;
- may be universal within organism and between organisms;
- regulation/regeneration due to reassignment of positional information.

Wolpert (1969). Jour Theor Biol 25: 1-47.

The French Flag Problem



slightly modified from: Wolpert (1968). In: Towards a Theoretical Biology (Waddington ed.)

Definition: Morphogen

Turing, 1952:

[...], the word being intended to convey the idea of a form producer. It is not intended to have any exact meaning, but is simply the kind of substance concerned in this theory. …diffuses and its distribution dictates the development of cells in the tissue

modern:

Molecule diffusing (or transported) from a localized source, which directly determines at least three territories of differing target gene expression over a large distance in a developmental field.

Turing (1952). Phil Trans R Soc B 237: 37.

Crick (1970). Nature 225: 420; Slack (1987). TIBS 12: 200; Neumann & Cohen (1997). BioEssays 19: 721; Gurdon & Bourillot (2001). Nature 413: 797.

Morphogenetic Field

A morphogenetic field is the region in which all cells have their position specified in respect to the same set of points or boundaries (Wolpert,1969).

- The general fate of a morphogenetic field is determined.
- However, the individual cells within the field are not committed, and the cells of the field can regulate their fates to make up for cells missing from the field.
- Moreover, if cells from one field are placed within another field, they can use the positional cues of their new location, even if they retain their organ-specific commitment.

Mosaic Development

Specification occurs autonomously, the embryo is constructed like a mosaic of independent, self-differentiating parts.

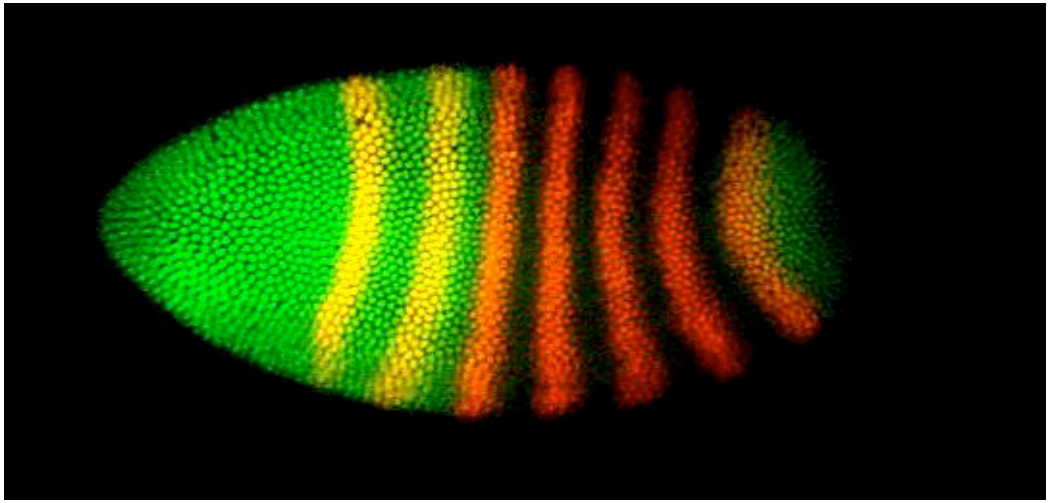
Removal of blastomers leads to loss of corresponding adult structures.

Regulative Development

Specification is conditional depending on neighboring cells.

If blastomere is removed from an early embryo, embryonic cells alter their fates so that the roles of the missing cells can be taken over.

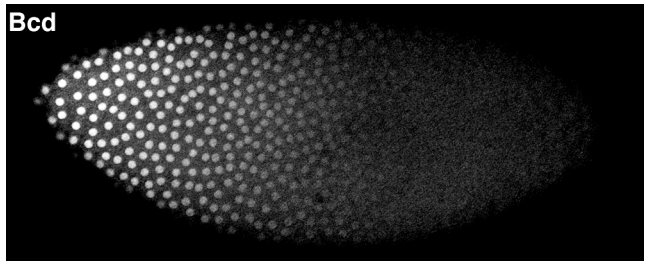
Expression of Segmentation Genes in *Drosophila*



0.5 mm long; fluorescently labeled with polyclonal antibodies.
The dorsal side is up and anterior
Is to the left; each dot is a nucleus.
No cell membranes are present.

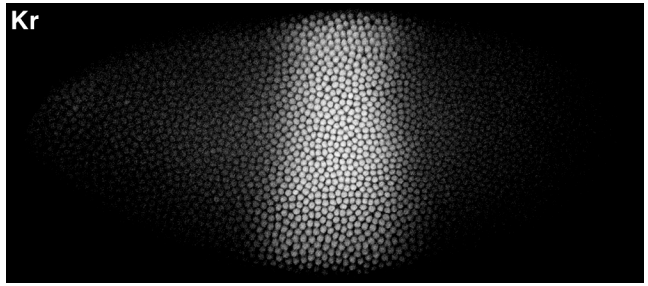
Three Classes of Segmentation Genes which Act in the Blastoderm

Maternal Coordinate Genes



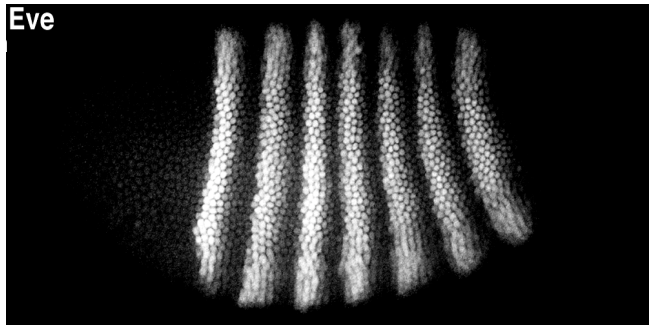
bicoid (bcd)
caudal (cad
hunchback

Gap Genes



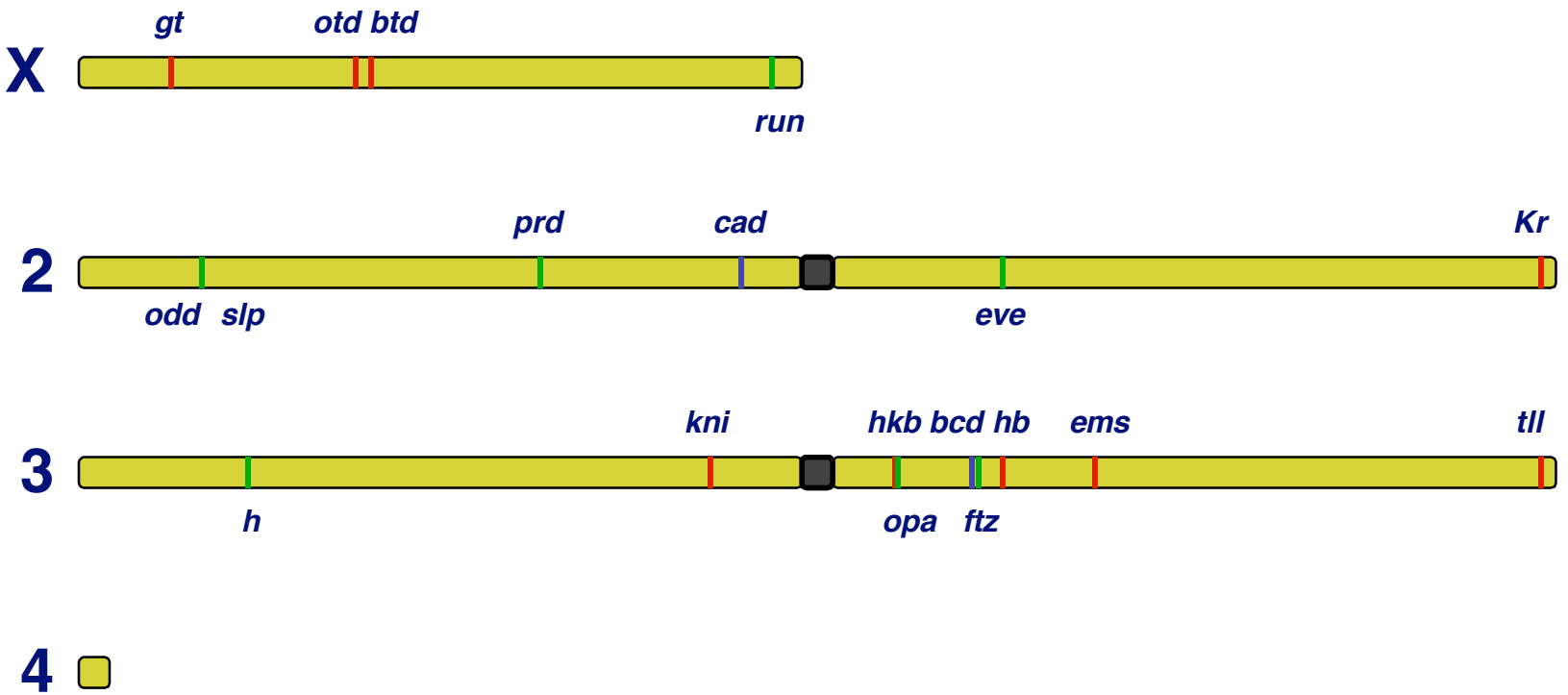
hunchback (hb)
Krüppel (Kr)
knirps (kni)
giant (gt)
tailless (tll)

Pair-Rule Genes



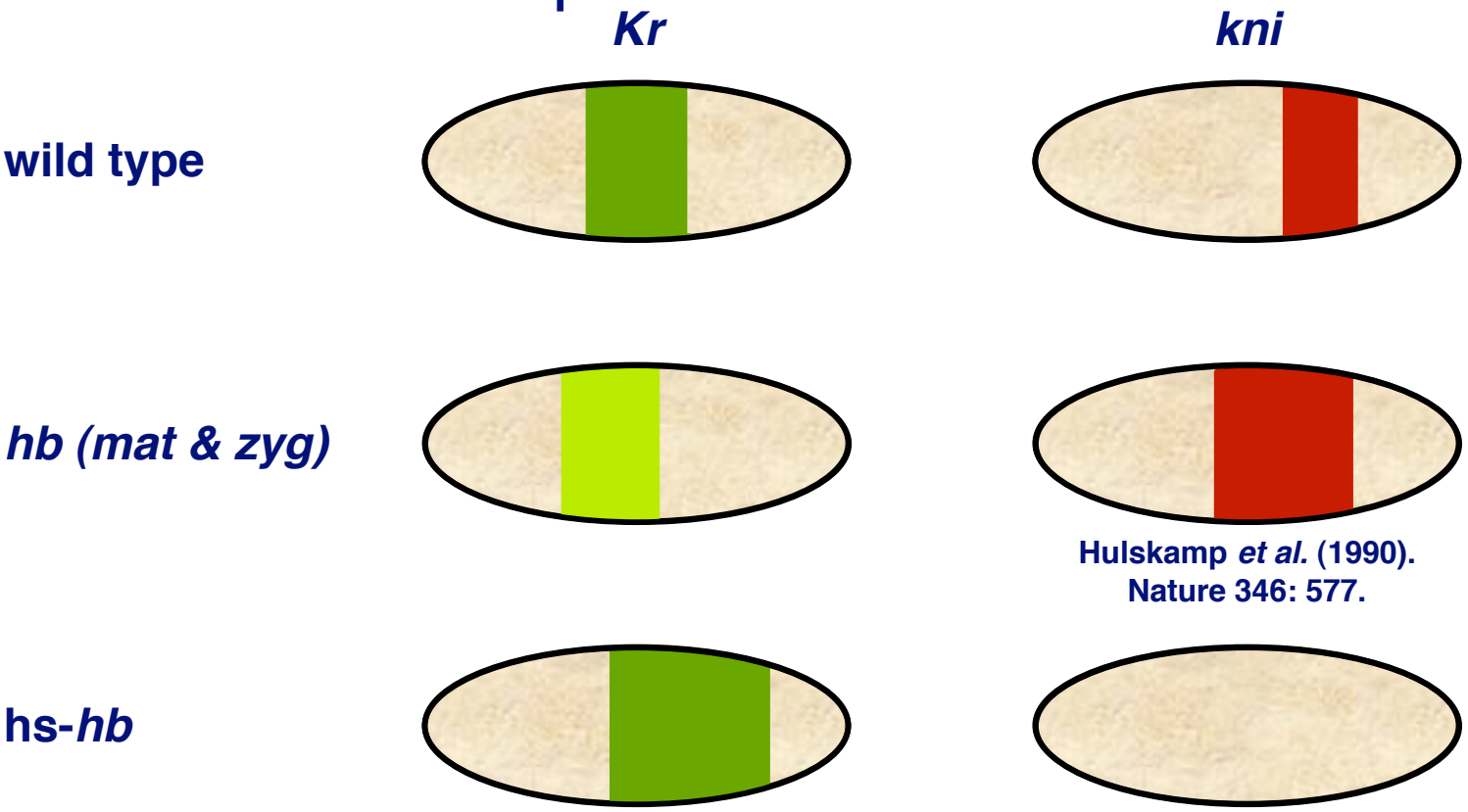
even-skipped (eve)
odd-skipped (odd)
hairy (h)
runt (run)
fushi-tarazu (ftz)
paired (prd)

Genomic Location of Segmentation Genes

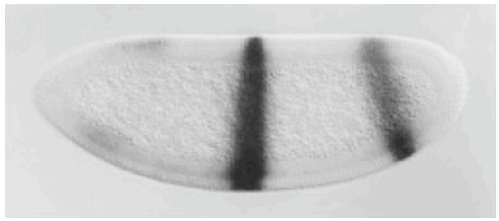


Pankratz & Jäckle (1993) / http://flybase.bio.indiana.edu

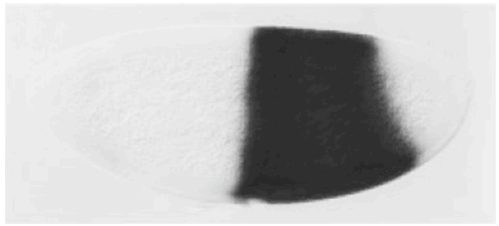
Interpretation of mutant expression patterns is difficult



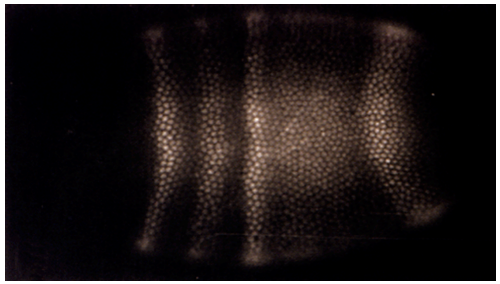
Regulatory non-equivalence of repoter constructs to endogenous gene



MSE3 *lac*Z in wild type embryo
(Small, Blair & Levine 1992, EMBO J. 11:4047)



MSE3 *lac*Z in *knirps* mutant embryo
(Small, Blair & Levine 1992, EMBO J. 11:4047)



Native *eve* in *knirps* mutant embryo
(Frasch and Levine 1987, G&D 1:981-995)

System Biology

- a new field in biology that aims at system-level understanding of biological systems.
- links the behaviors of molecules to system characteristics and functions.
- constitutes a major multi-disciplinary research effort that will enable us to understand biological systems as systems.
- requires collective efforts in four key areas: (1) genomics and other molecular biology research, (2)computational studies, such as simulation, bioinformatics, and software tools, (3) analysis of dynamics of the system, and (4) technologies for highprecision, comprehensive measurements.

Hiroaki Kitano "Systems Biology: Toward System-level Understanding of Biological Systems"

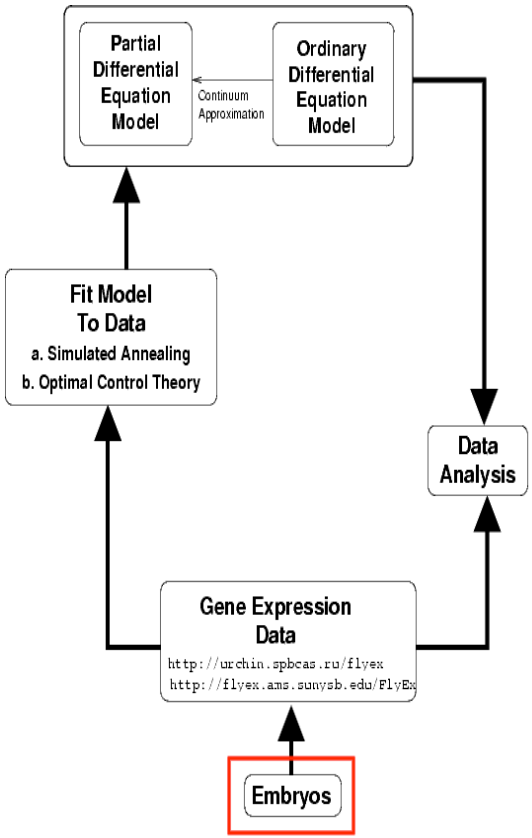
The Gene Circuit Approach

A data-driven mathematical modeling method whose main aim is to extract information about dynamical regulatory interactions between transcription factors from given gene expression patterns. This is achieved in four steps:

- Obtain gene expression data
- Formulate a theoretical model
- Perform optimization to fit model to data
- Learn some new biology

Mjolsness et al., 1991; Reinitz etal., 1995; Reinitz and Sharp, 1995; Reinitz et al., 1998

System Biology of Segment Determination



Model Parameters

Cost Functions

Shifts,
Positional Errors,
Strip Asymmetry

Database Records,
Numerical Text Files

Images

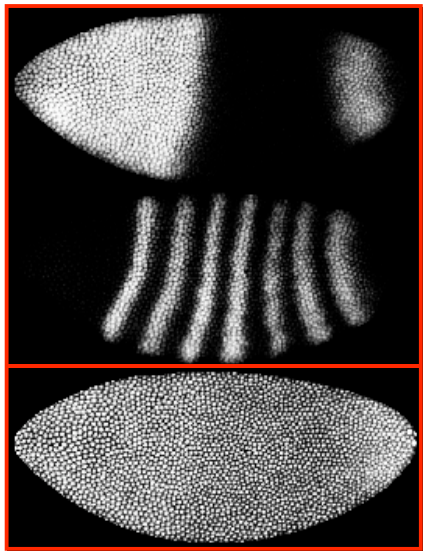
Nature, v 430, 15
July 2004: 368-371

14 Segmentation Genes Are Expressed in the Blastoderm. Each Has a Distinct Pattern of Expression.

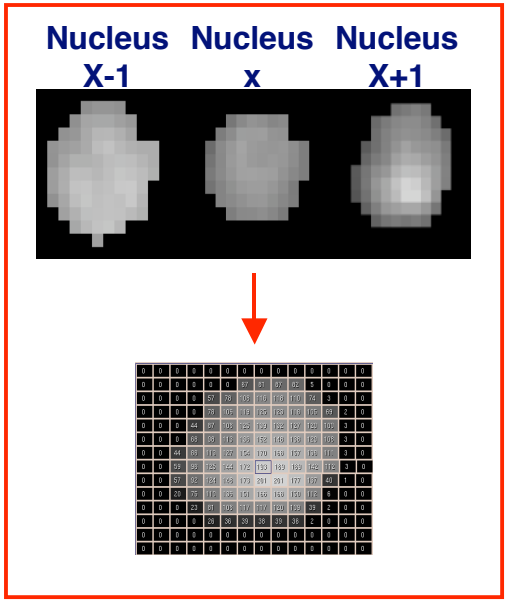
Acquisition of Quantitative Gene Expression Data

- Segment embryos
- Classify embryos into time classes
- Register the data
- Remove the background
- Average the data

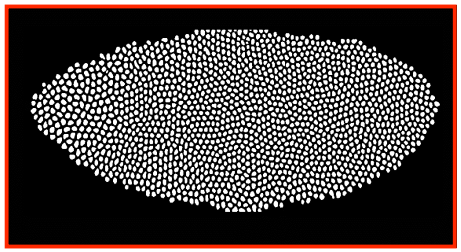
Image segmentation



Blob extractor
Grey scale intensity values of pixels are grouped based on mask elements, each representing a single nucleus. Each group of values is then averaged to yield the fluorescence intensity values assigned to a specific nucleus.



Nuclear Mask

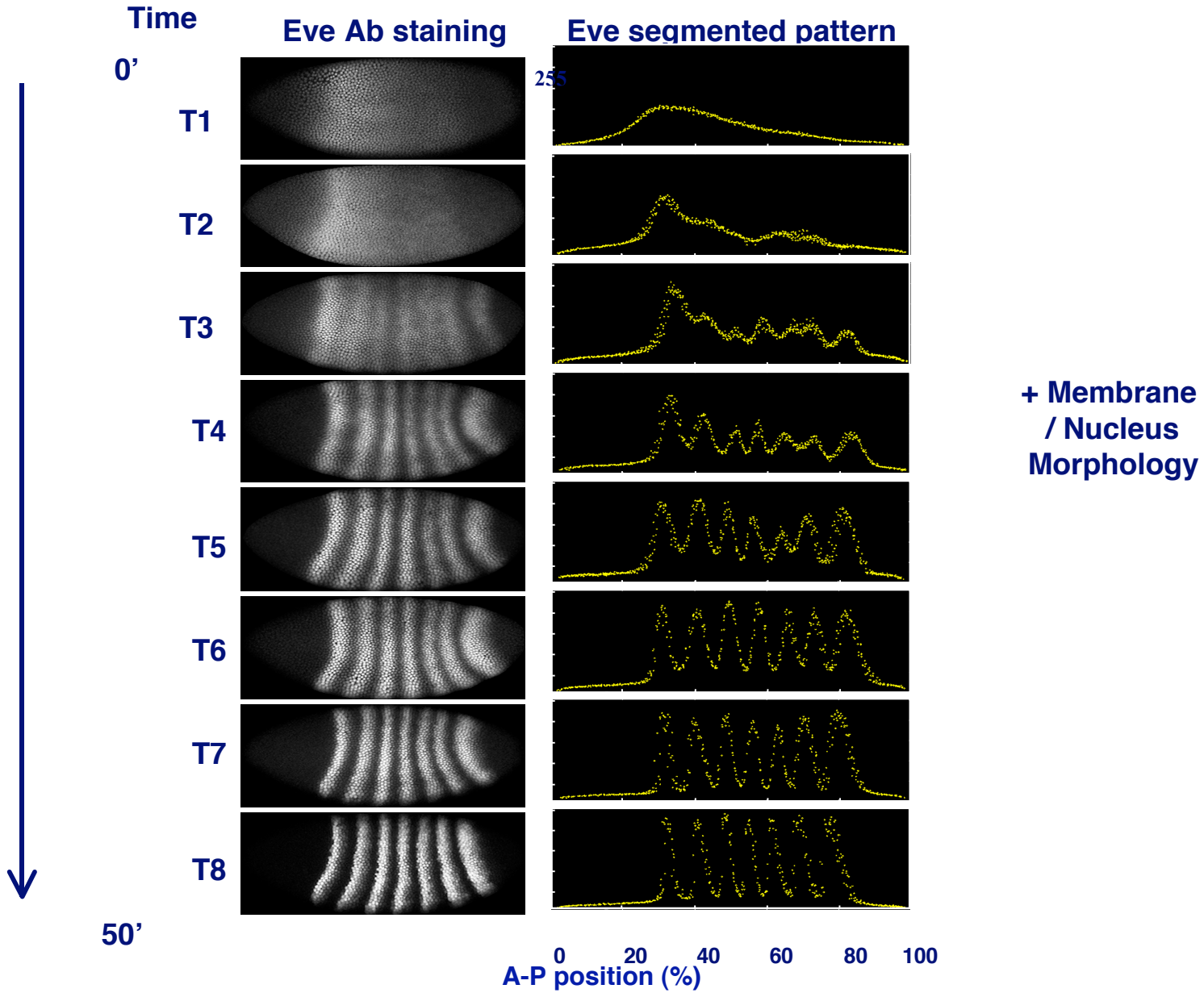


Shape Analysis
The average x and y location of all the pixels in a group (centroid) is calculated thus defining the position of the nucleus.

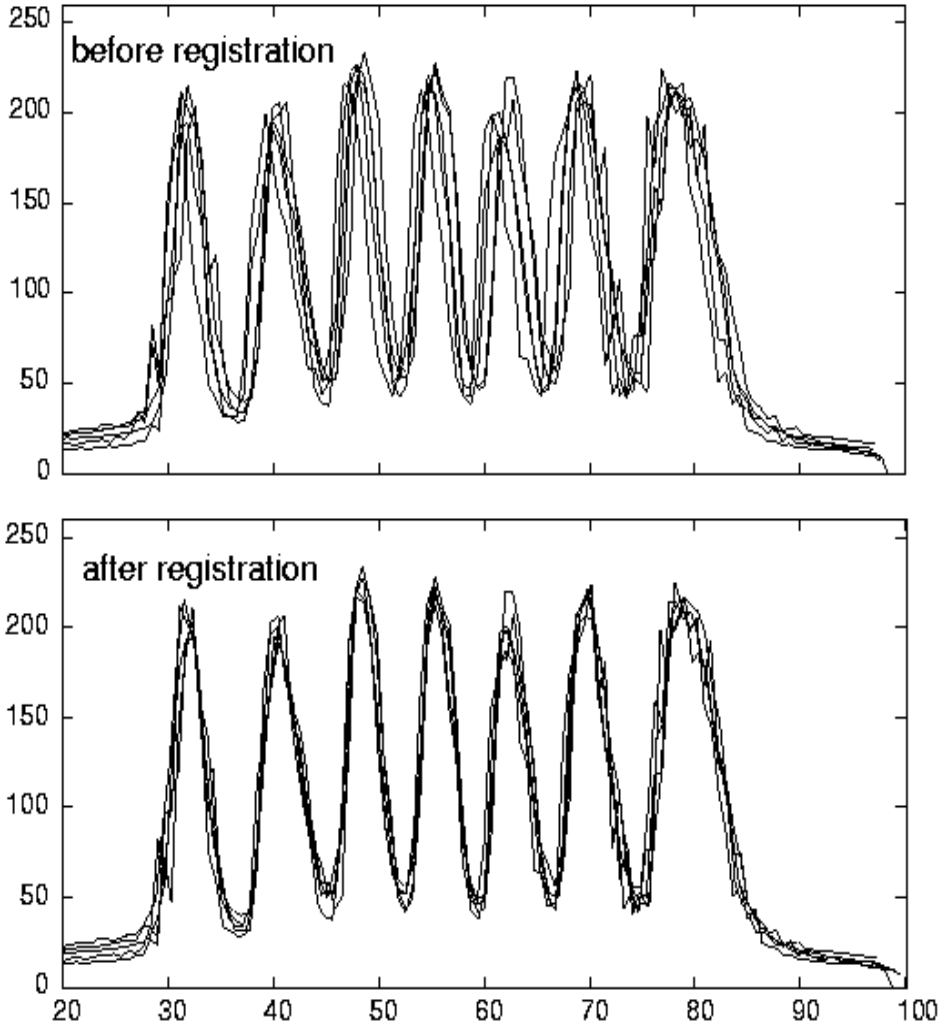
	Nuclear Coordinates		Fluorescence intensities	
0	3.67346	41.5401	53.6951	25.939
1	4.05099	45.3917	53.8919	27.0811
2	4.00196	37.3767	54.7	22.85
3	5.01298	39.8973	59.4688	26.5625
4	5.07766	52.1475	58.8267	29.04
5	5.44521	43.4772	62.25	30.5125
6	5.52642	47.3609	56.7215	28.9494
7	5.78564	36.9171	56.8689	26.2623
8	6.26299	50.786	64.3521	30.6056
9	6.62735	40.1422	59.3086	32.0123
10	6.69946	54.7947	60.0795	29.5227

Classify embryos into time classes

Myasnikova et al., Bioinformatics 17 : 3 – 12



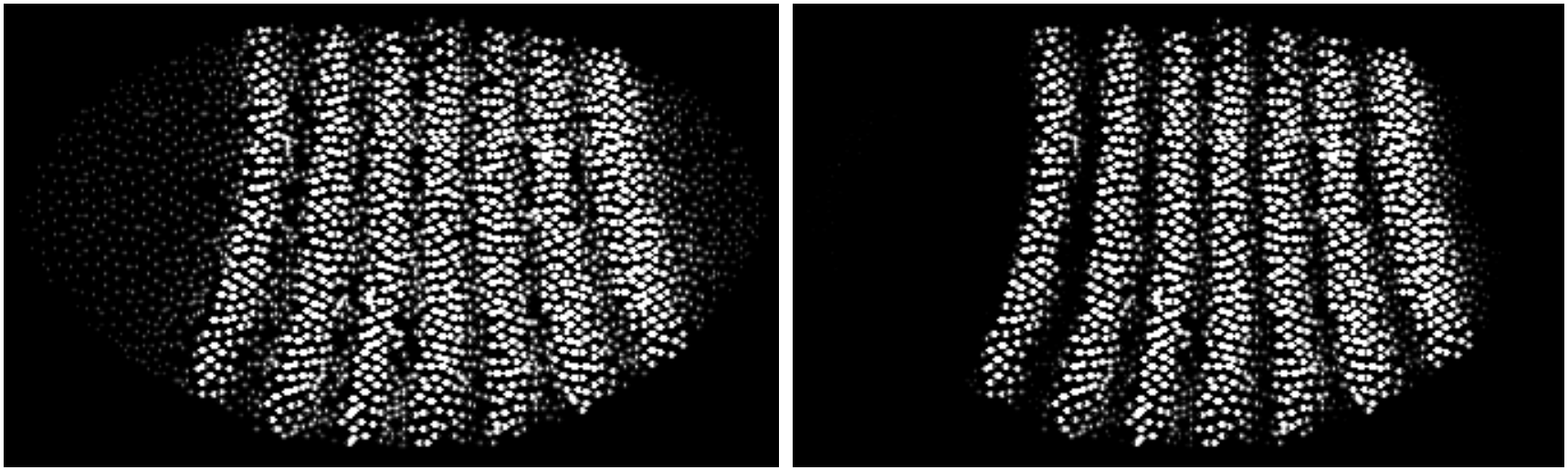
Registration of 5 *eve* Expression Patterns



Myasnikova et al.(1999)
Proc. ISMB99,
pp. 195-201;

Myasnikova et al. (2001)
Bioinformatics 17:3-12

Background Removal



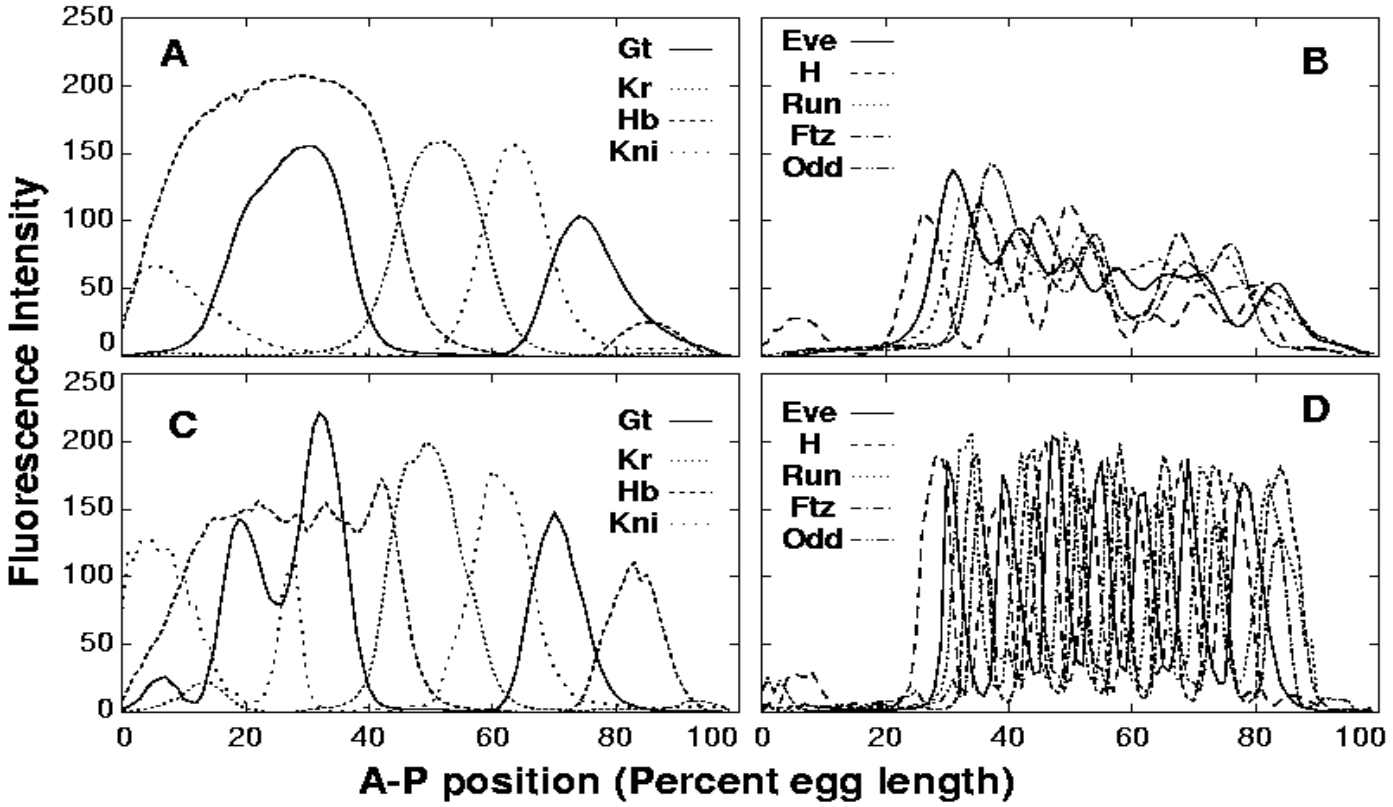
$$B(x,y) = a\left(x^2 + (100 - y)^2\right) + b(x + (100 - y)) + c$$

$$z_{norm,i} = 255 * \frac{z_i - B(x_i, y_i)}{255 - B(x_i, y_i)}$$

Construction of the Integrated Pattern of Gene Expression at Cellular Resolution

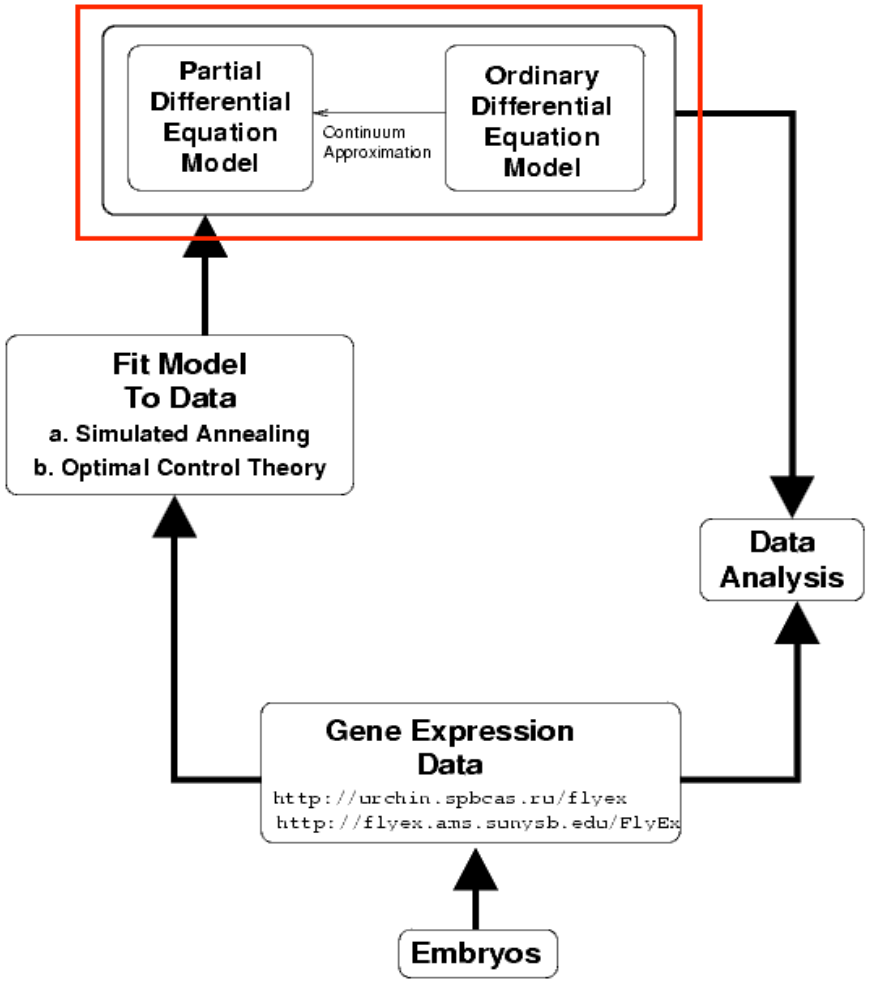
1. Group the coordinates of nuclei in each pattern in the registered dataset along the *x* axis into *R* intervals;
2. Within each interval, compute the average value of protein concentration over all the embryos in the time class;
3. Choose *R=100* because of the fact that a single nucleus is very close to 1% egg length in diameter.

Integrated Patterns of Segmentation Gene Expression



Myasnikova et al (2001), Bioinformatics 17:3-12

System Biology of Segment Determination



Model Parameters

Cost Functions

Shifts,
Positional Errors,
Strip Asymmetry

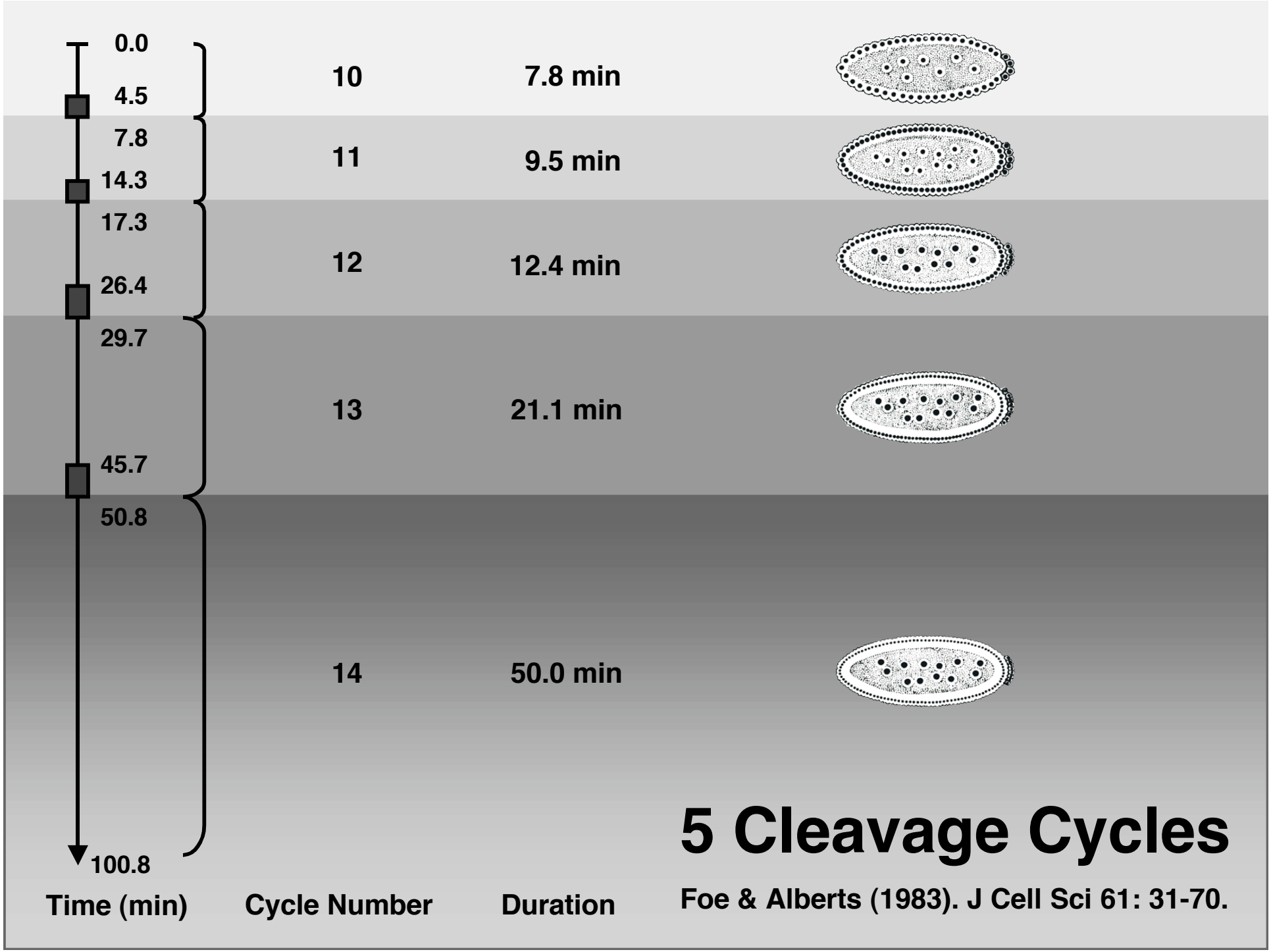
Database Records,
Numerical Text Files

Images

Why are blastoderm models solvable?

The blastoderm is unusually tractable for modeling a mechanism of determination of a morphogenetic field as

- ✓ **gene expression is observable,**
- ✓ **only a small number of genes are expressed from the zygote,**
- ✓ **the blastoderm is syncitium of nuclei and cell-cell signaling can be neglected,**
- ✓ **during the blastoderm stage gene regulation is uncoupled from other developmental processes.**


0.0
4.5
7.8
14.3
17.3
26.4
29.7
45.7
50.8
100.8
10
7.8 min
11
9.5 min
12
12.4 min
13
21.1 min
14
50.0 min
Time (min)
Cycle Number
Duration
5 Cleavage Cycles
Foe & Alberts (1983). J Cell Sci 61: 31-70.

Hybrid Dynamical System:

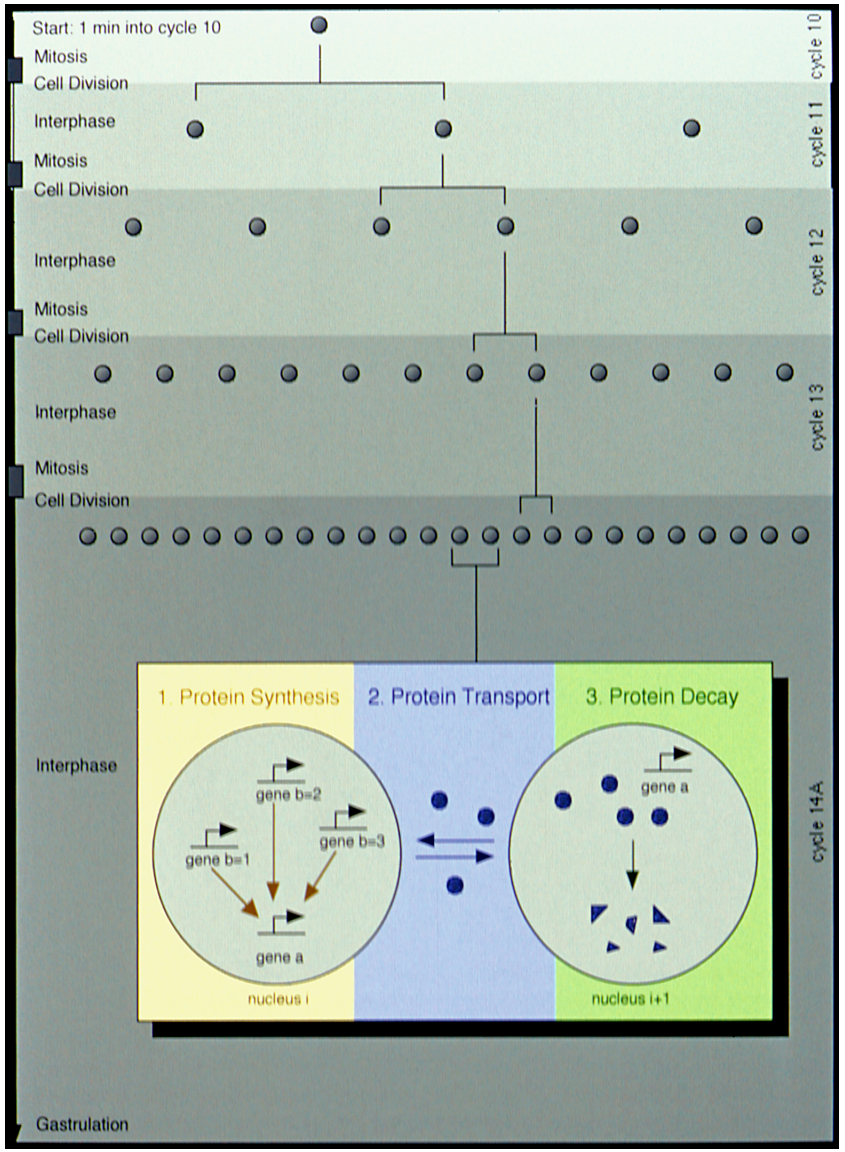
Two kinds of continuous dynamics:

1. **Interphase (synthesis)**
2. **Mitosis (no synthesis)**

and a discrete transition:

3. Cleavage.

Mjolsness, Sharp & Reinitz (1991), J. Theor. Biol. 152: 429-453



Synthesis

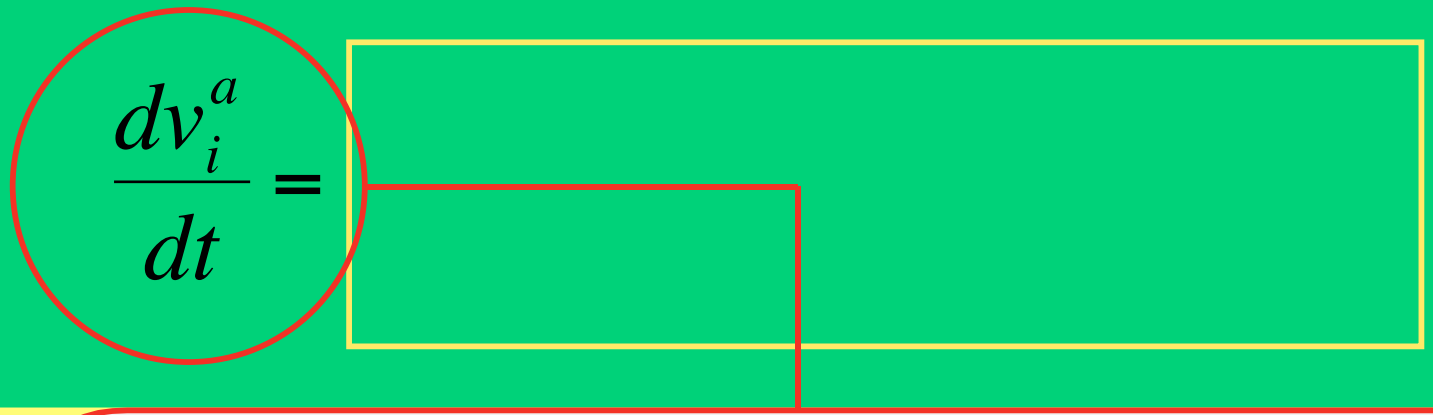
$$\frac{dv_i^a}{dt} = R_a g_a \left(\sum_{b=1}^{N} T^{ab} v_i^b + m^a v_i^{bcd} + h_a \right)$$

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

$$-\lambda_a v_i^a$$



Differential Equations:

Each equation describes the rate of change in concentration of each segmentation gene product in each nucleus over time.

In the gap gene circuit there are

58 nuclei x 6 genes = 348 equations

Synthesis

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

Regulatory Input on Gene a:

The regulatory contribution of each gene b is proportional to its concentration.

Synthesis

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

Regulatory Input on Gene a:

The regulatory contribution of each gene b is proportional to its concentration.

bcd is an external input to the system.

Synthesis

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

Regulatory Input on Gene a:

The regulatory contribution of each gene b is proportional to its concentration.

bcd is an external input to the system.

h represents a promoter threshold, i.e the expression state of gene a in the absence of all other gene products in the model.

Synthesis

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

Regulatory Input on Gene a:

The regulatory contribution of each gene b is proportional to its concentration.

bcd is an external input to the system.

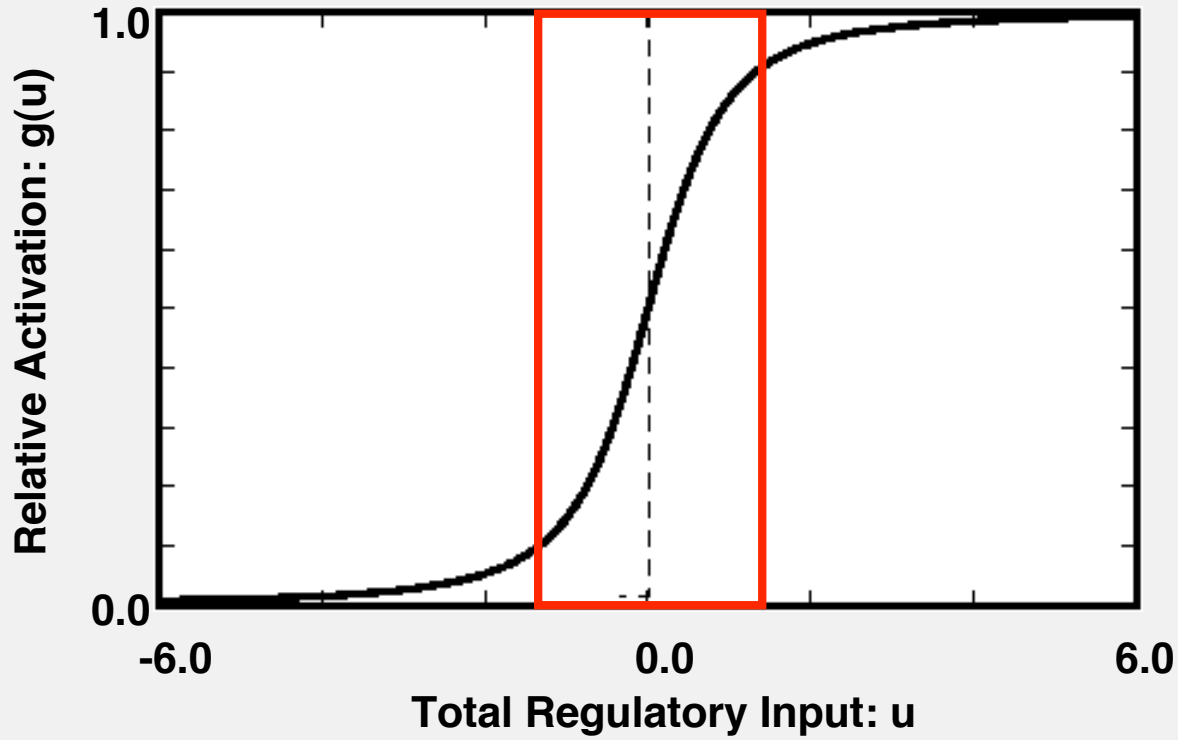
h represents a promoter threshold, i.e the expression state of gene a in the absence of all other gene products in the model.

Regulatory contributions are additive.

Synthesis

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

The regulation-expression function g(u):



Synthesis

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

Genetic Interconnectivity Matrix (T):

Gene a \ b	1	2	⋯	N
1	T^{11}	T^{12}	⋯	T^{1N}
2	T^{21}	T^{22}	⋯	T^{2N}
⋮	⋮	⋮		⋮
N	T^{N1}	T^{N2}	⋯	T^{NN}

T parameters:

positive: activation
negative: repression
zero: no interaction

Synthesis

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

Transport

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

Protein Transport:

Proportional to concentration differences between both neighboring nuclei.

Synthesis

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

Transport

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

Diffusion Schedule:

Protein transport increases as nuclei divide and the distance between nuclei decreases.

Synthesis

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

Transport

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

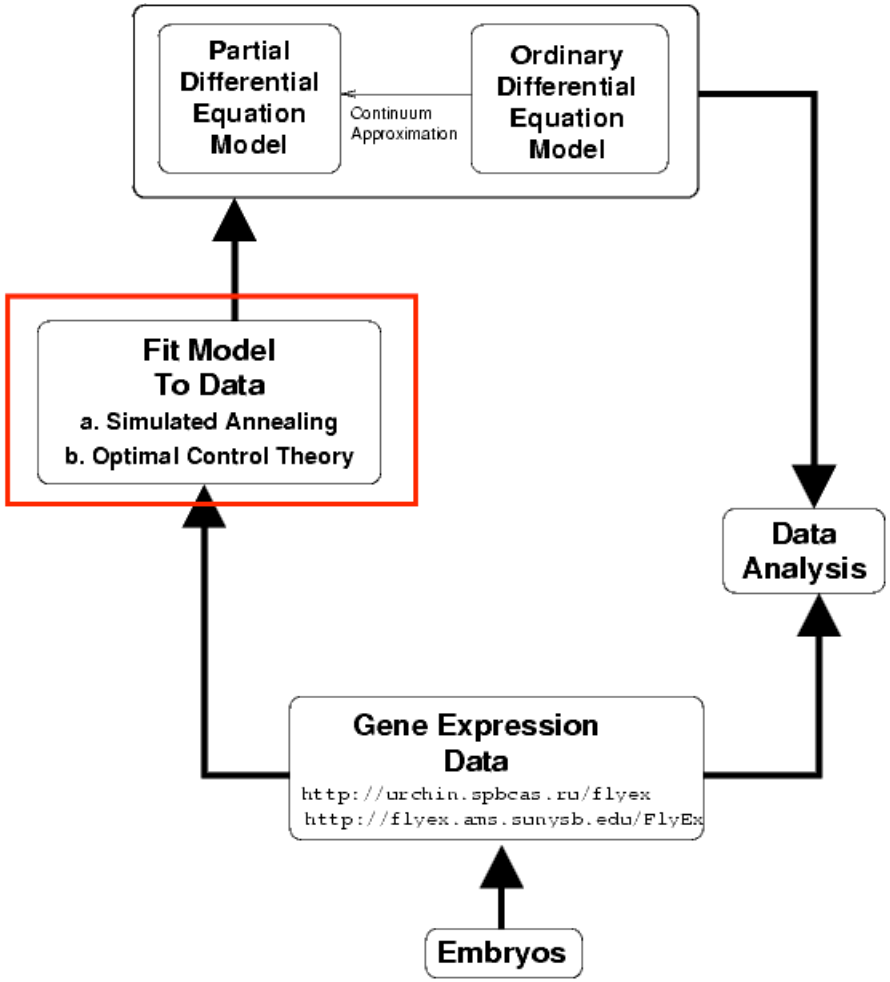
Decay

$$-\lambda_a v_i^a$$

Protein Decay:

Proportional to protein concentration.

System Biology of Segment Determination



Model Parameters

Cost Functions

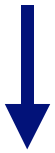
Shifts,
Positional Errors,
Strip Asymmetry

Database Records,
Numerical Text Files

Images

Fits to data

INPUT: gene expression patterns, random parameters



Change parameters until model produces patterns that are as similar as possible to gene expression data

OUTPUT: gene circuit (particular set of parameters)

Fit to Data

We seek parameters for interphase equations that minimize

$$E = \sum_{\substack{\text{all } a,\ i \\ \text{and } t \\ \text{for which} \\ \text{data exist}}} \left(v_i^a(t)_{\mathrm{model}} - v_i^a(t)_{data} \right)^2$$

We minimize E using the method of simulated annealing.

Simulated Annealing

- **Advantage: The method will yield the *global minimum* of *E.***
- **Cost: Computationally intensive.**
  - **serial simulated annealing: 12 hrs - 52 days on a 2 Ghz Pentium P4**
  - **parallel simulated annealing: 4 hrs - 5 days on 10 1.5Ghz AMD processors**

Simulated Annealing

Metropolis et al.(1953); Kirkpatrick et al. (1983)

1. **Compute $E = E_{old}$ from the variables x_i .**
2. **Make a change in one (or more) of the x_i (this is referred to as a "move".**
3. **Compute $E = E_{new}$ from the newly generated set of x_i .**
4. **Compute $\exp((E_{old} - E_{new}) / T)$.**
5. **If the above quantity is bigger than a random number between zero and one, keep the new x_i's ("accept the move"). Otherwise, restore the old x_i's ("reject the move").**
6. **Repeat while allowing T to decrease slowly from a large value to zero. Typically this entails 10^5 to 10^9 iterations.**