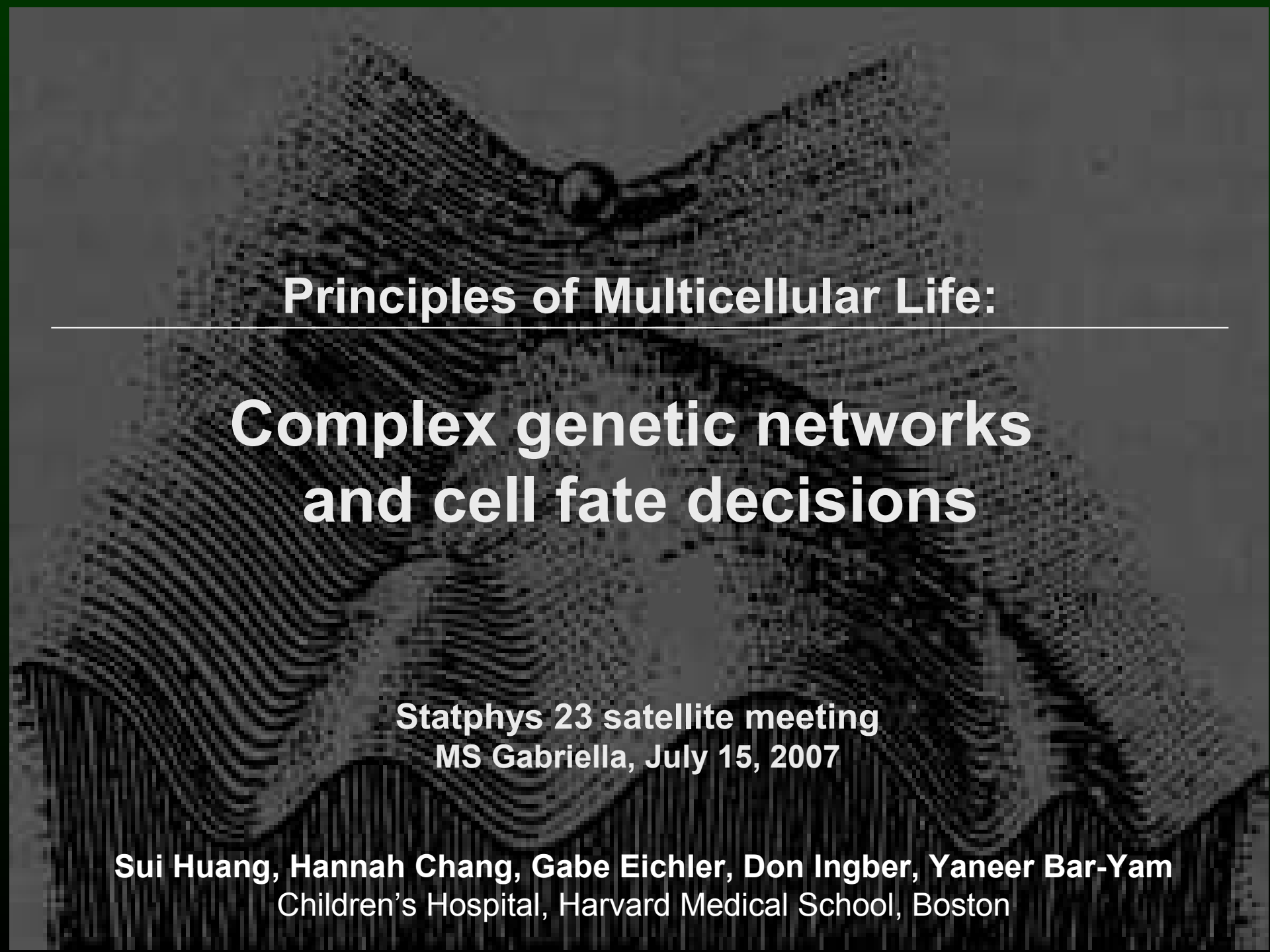


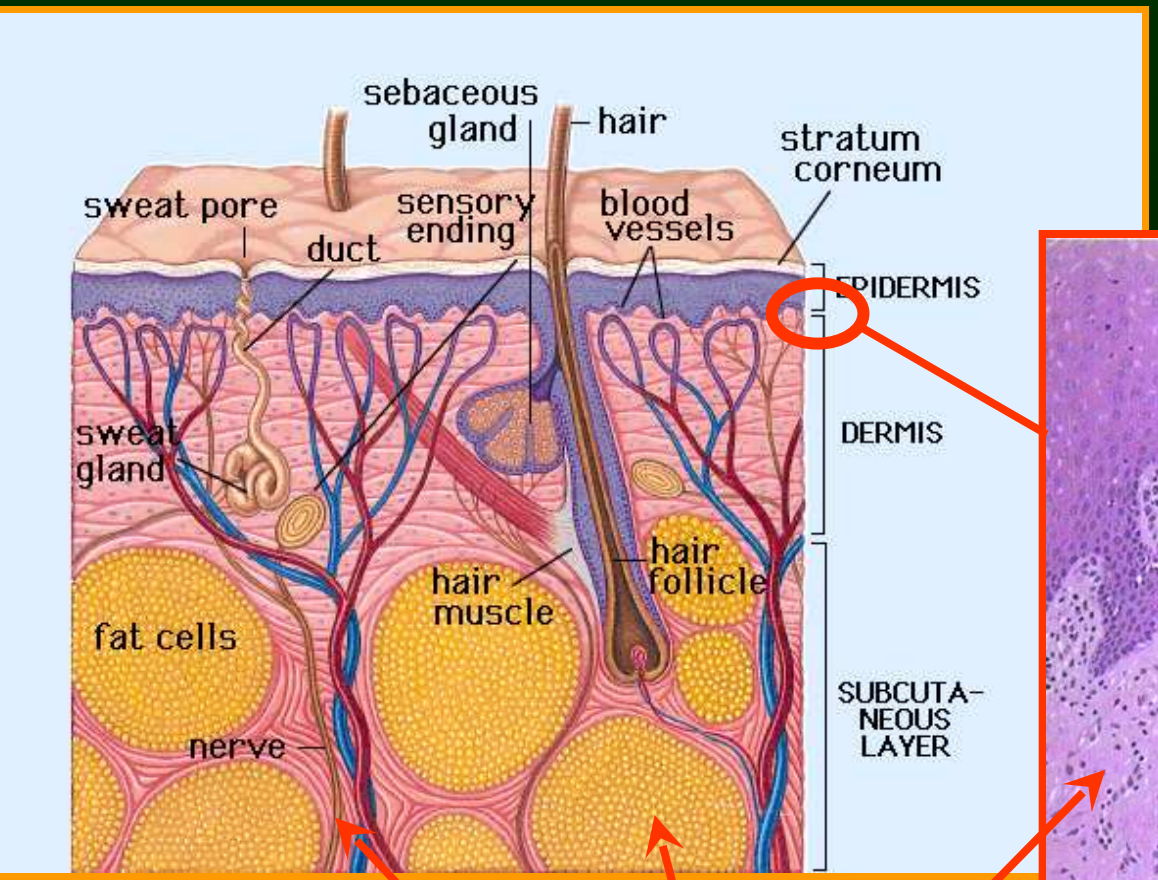
Principles of Multicellular Life:

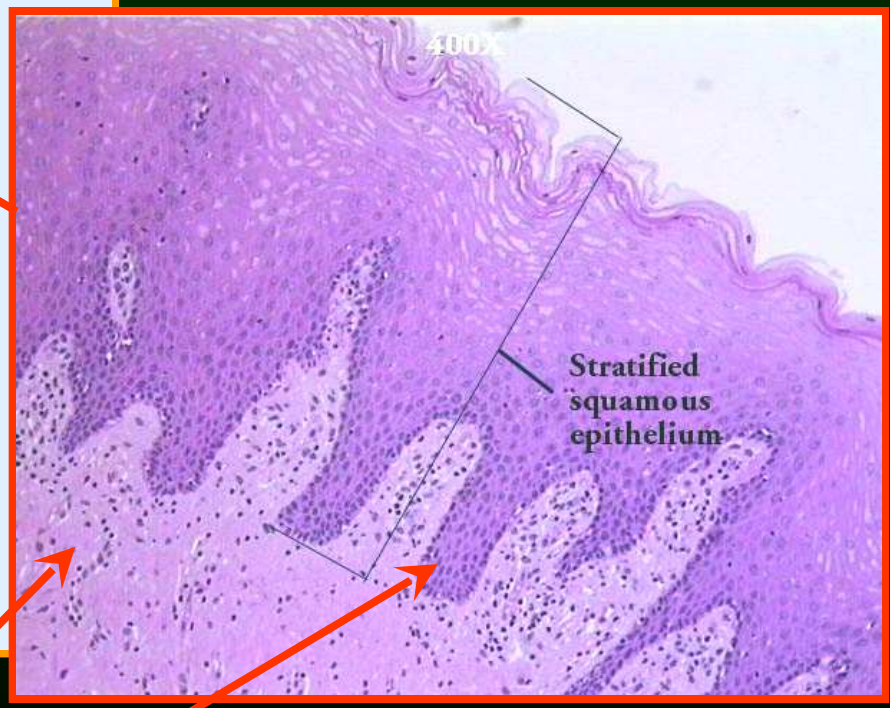
Complex genetic networks and cell fate decisions

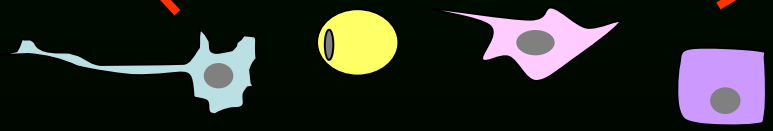
Statphys 23 satellite meeting
MS Gabriella, July 15, 2007

Sui Huang, Hannah Chang, Gabe Eichler, Don Ingber, Yaneer Bar-Yam
Children's Hospital, Harvard Medical School, Boston

Variety of cell types in metazoan tissue

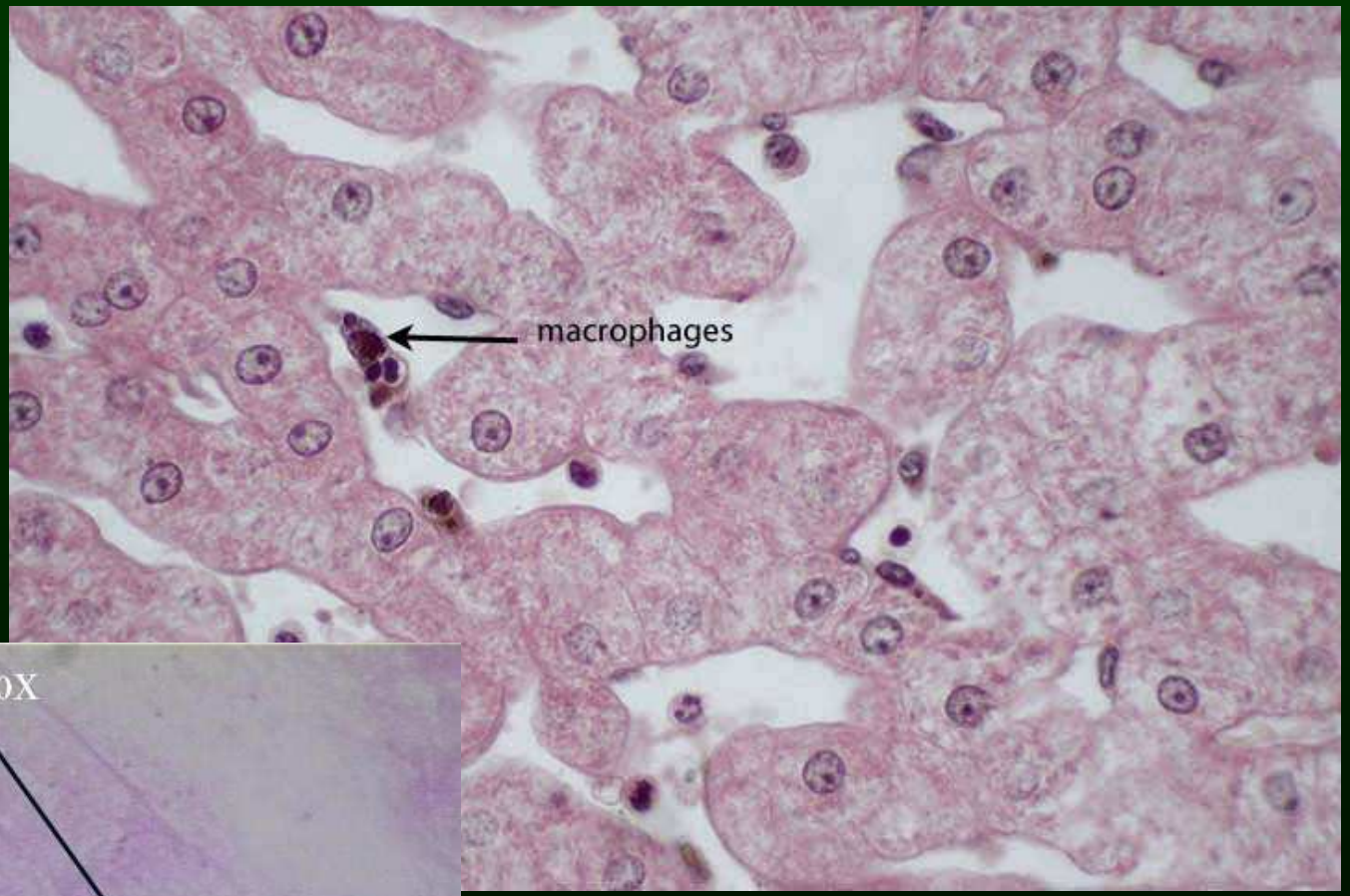


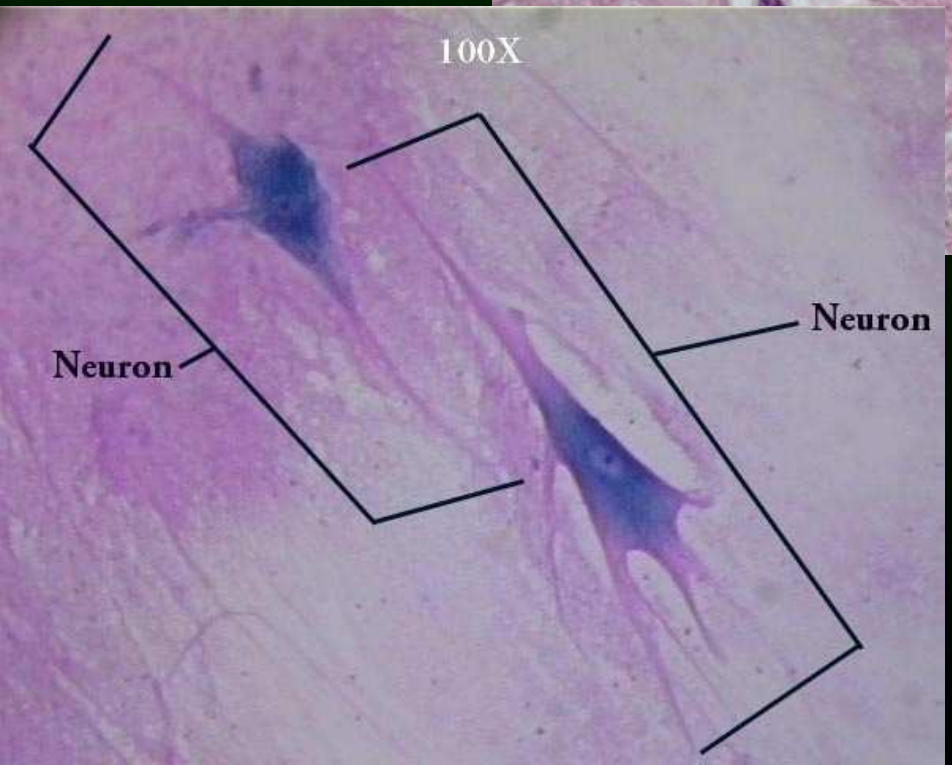




Every cell has the same genome = same genetic instructions

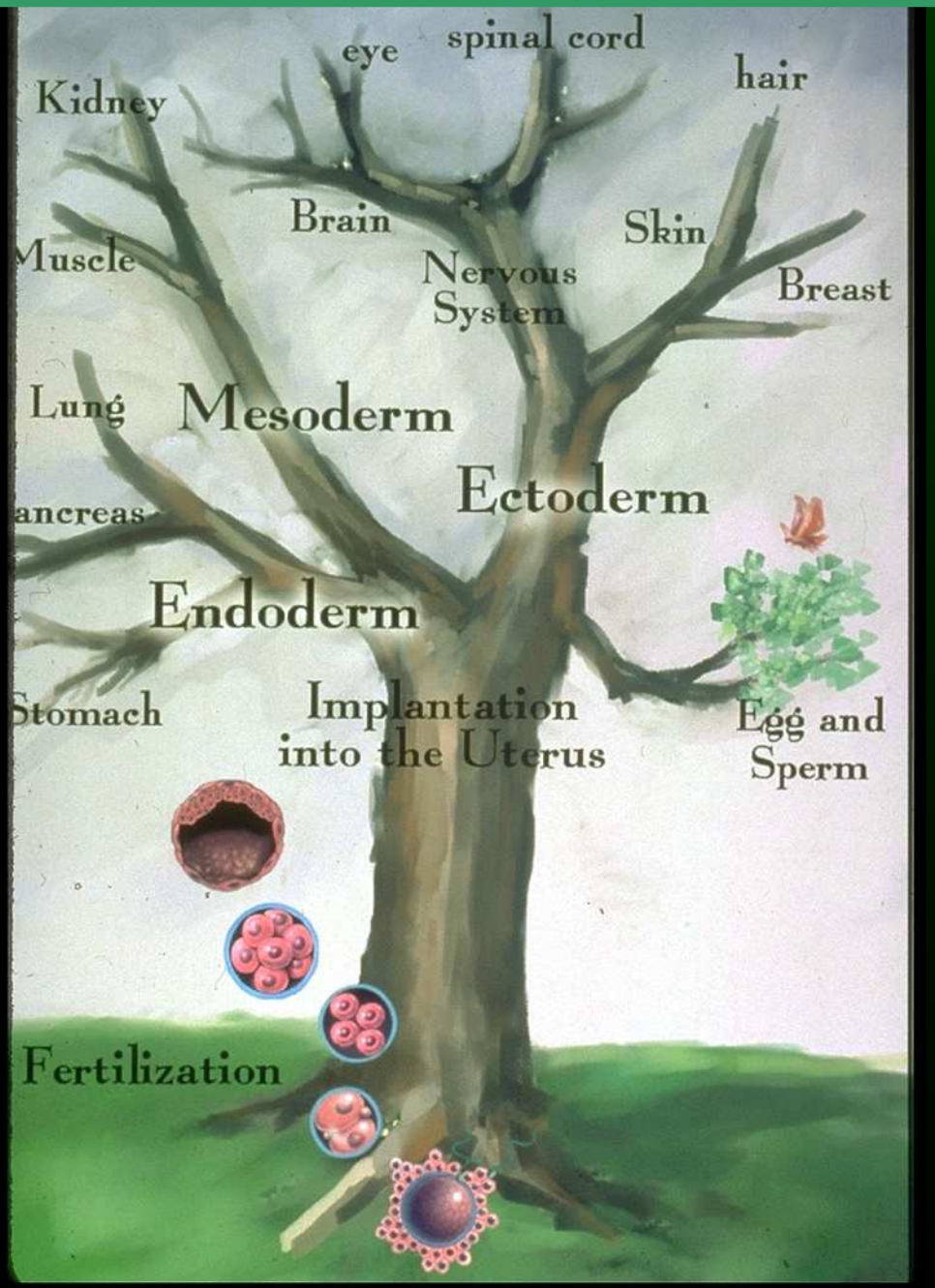
For non-biologists: Some fundamental features



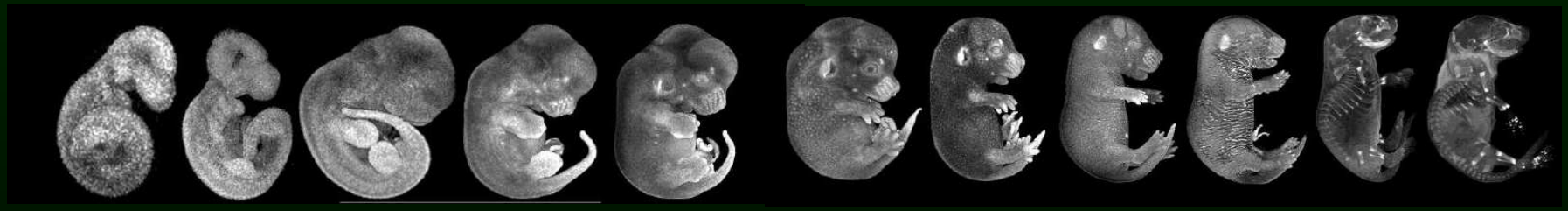


***Cell types*:**
Stable, discretely distinct phenotypic entities – much like different species !

Tree of Embryonic development



Arrow of time of development

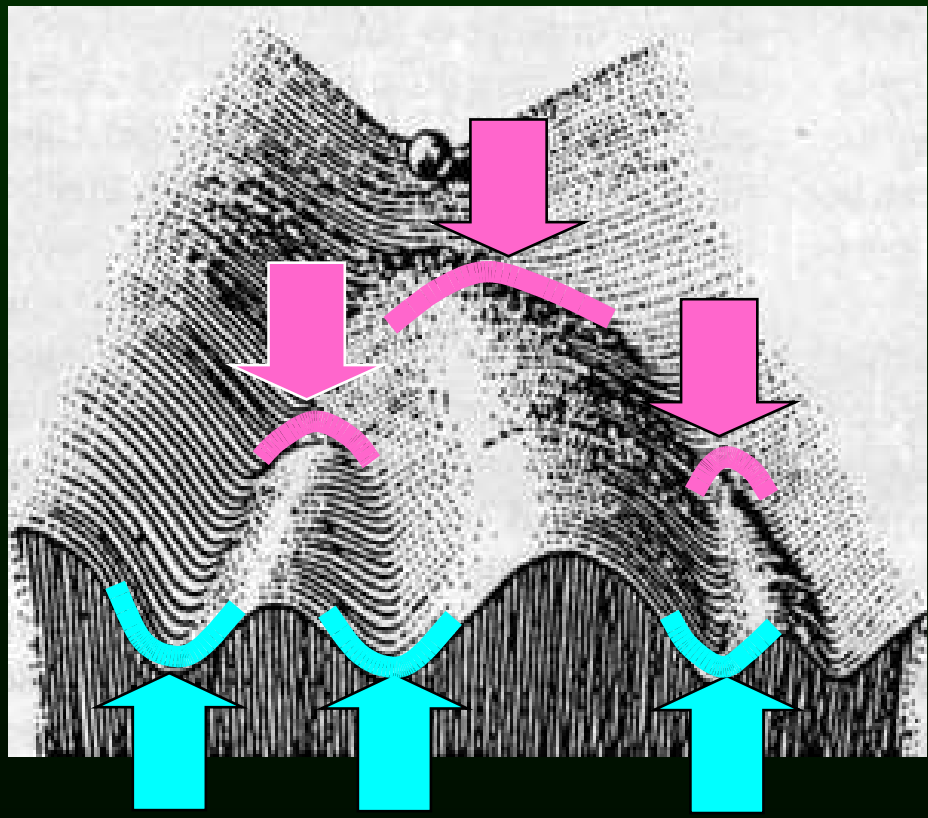


time

Waddington's "Epigenetic landscape"

C.F. Waddington, 1940s:

- *Cells "switch between **well-recognisable** types".*
- *"**Intermediates** are rare"*



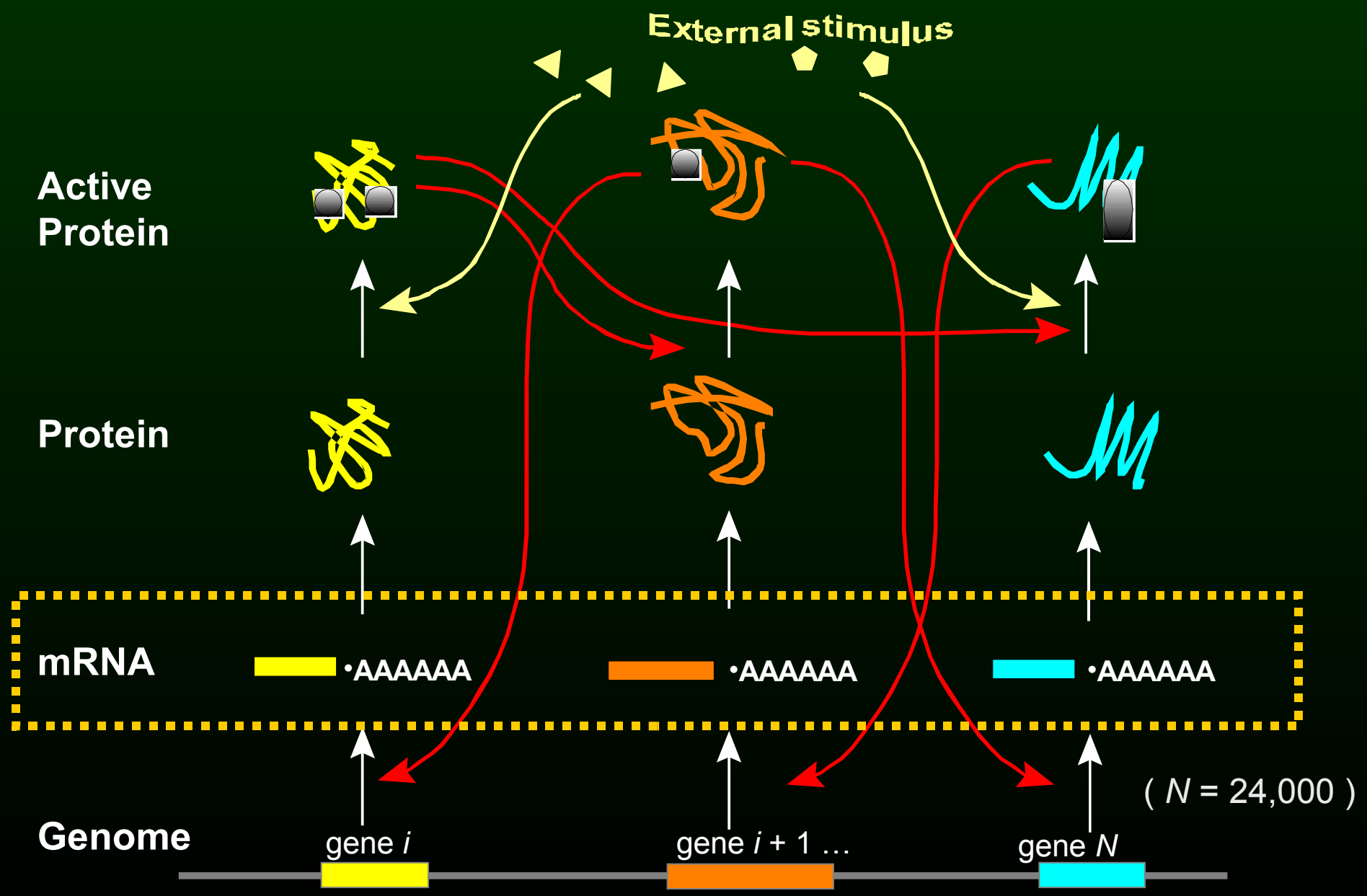
Two central questions:

What is the molecular / formal basis of this landscape?

- ***Valleys*: discrete stable states = cell fates / cell types**
- ***Hills:* unstable states = decision points**

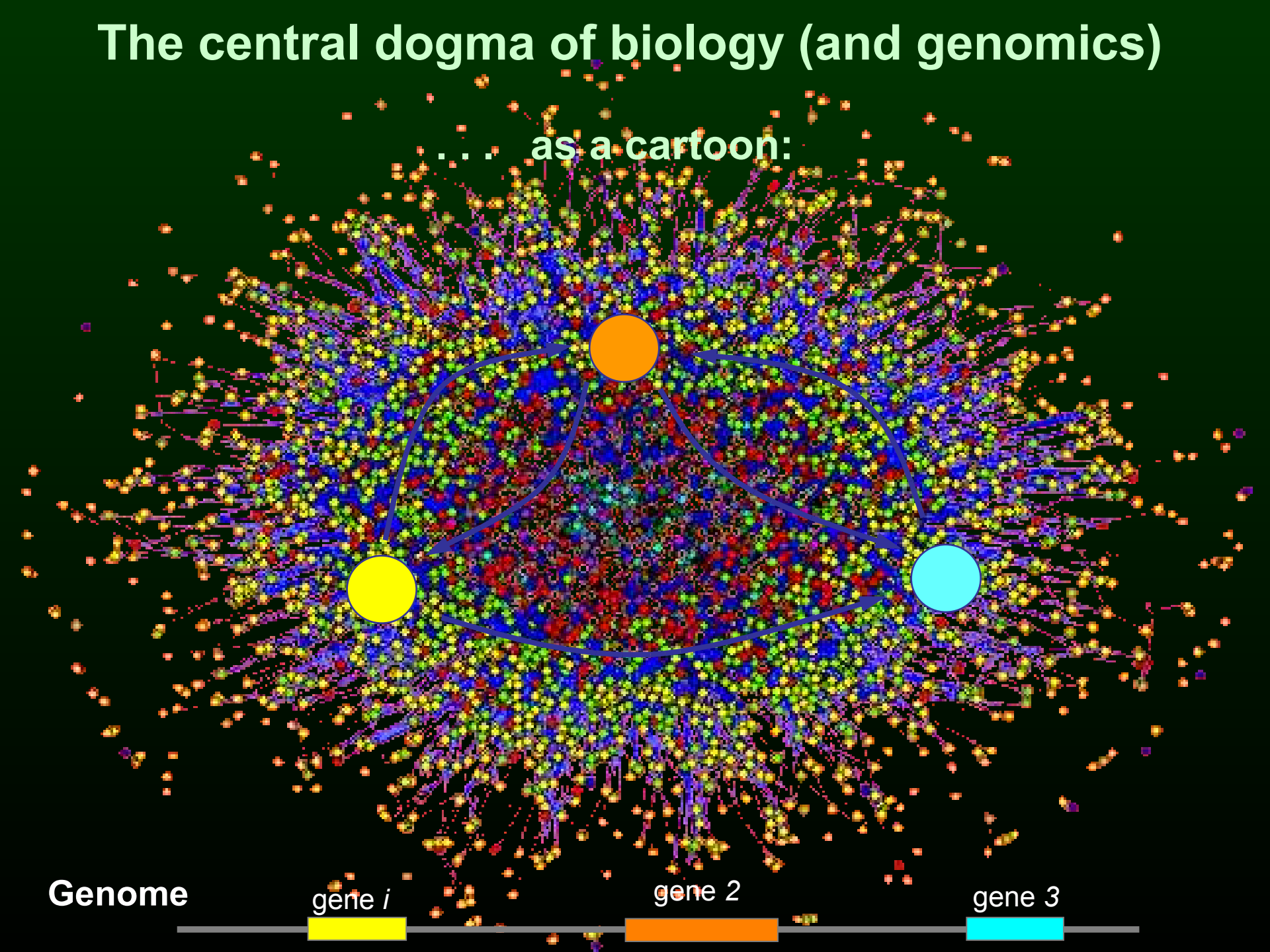
The molecular machinery underneath

The central dogma of biology (and genomics)

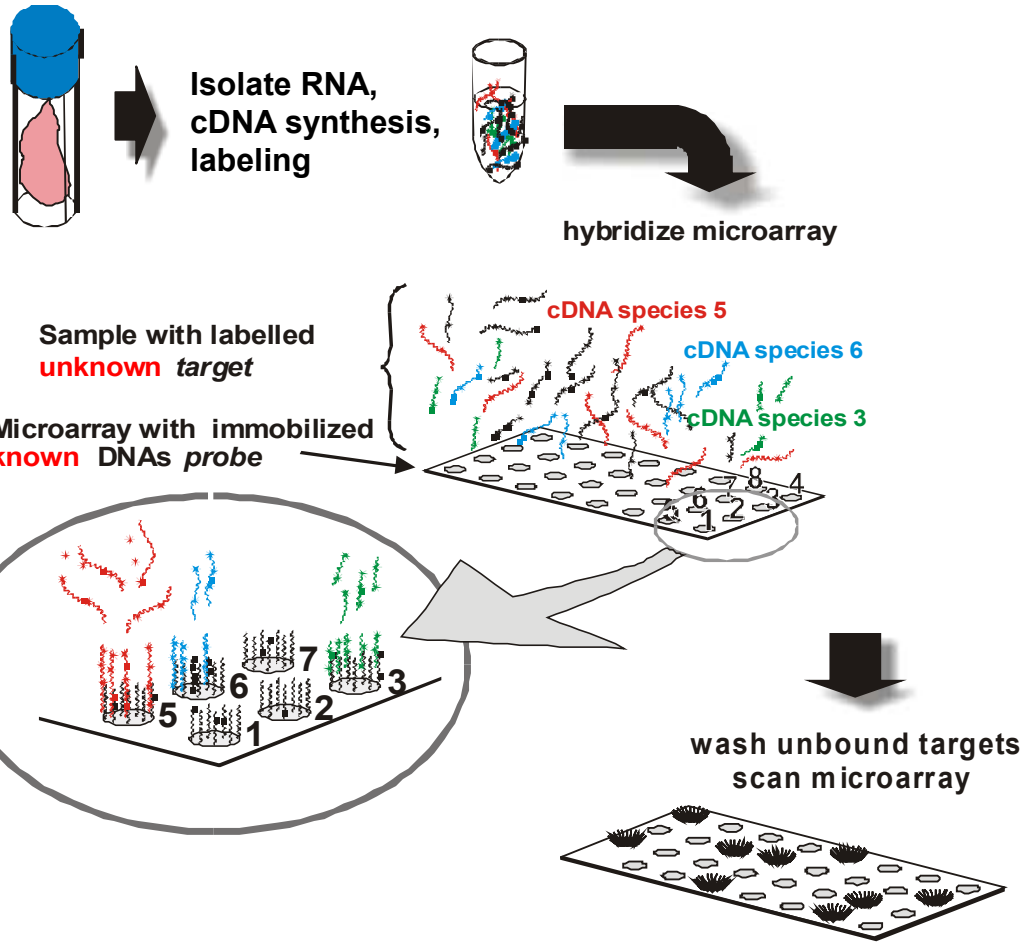


The central dogma of biology (and genomics)

. . . as a cartoon:

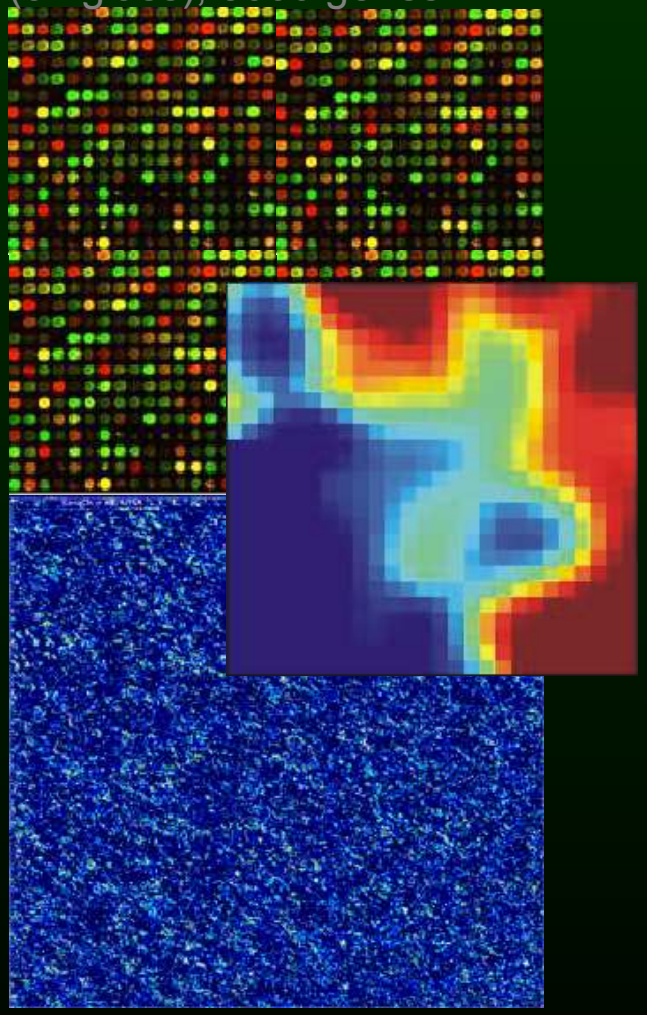


Principle of DNA microarray expression profiling



	Sample A	Sample B
Gene 1	164.34	86.41
Gene 2	12.32	42.31
Gene 3	456.77	416.57
Gene 4	0.02	0.06
Gene 5	214.98	233.18
Gene 6	114.61	112.6
Gene 7	27.54	87.3

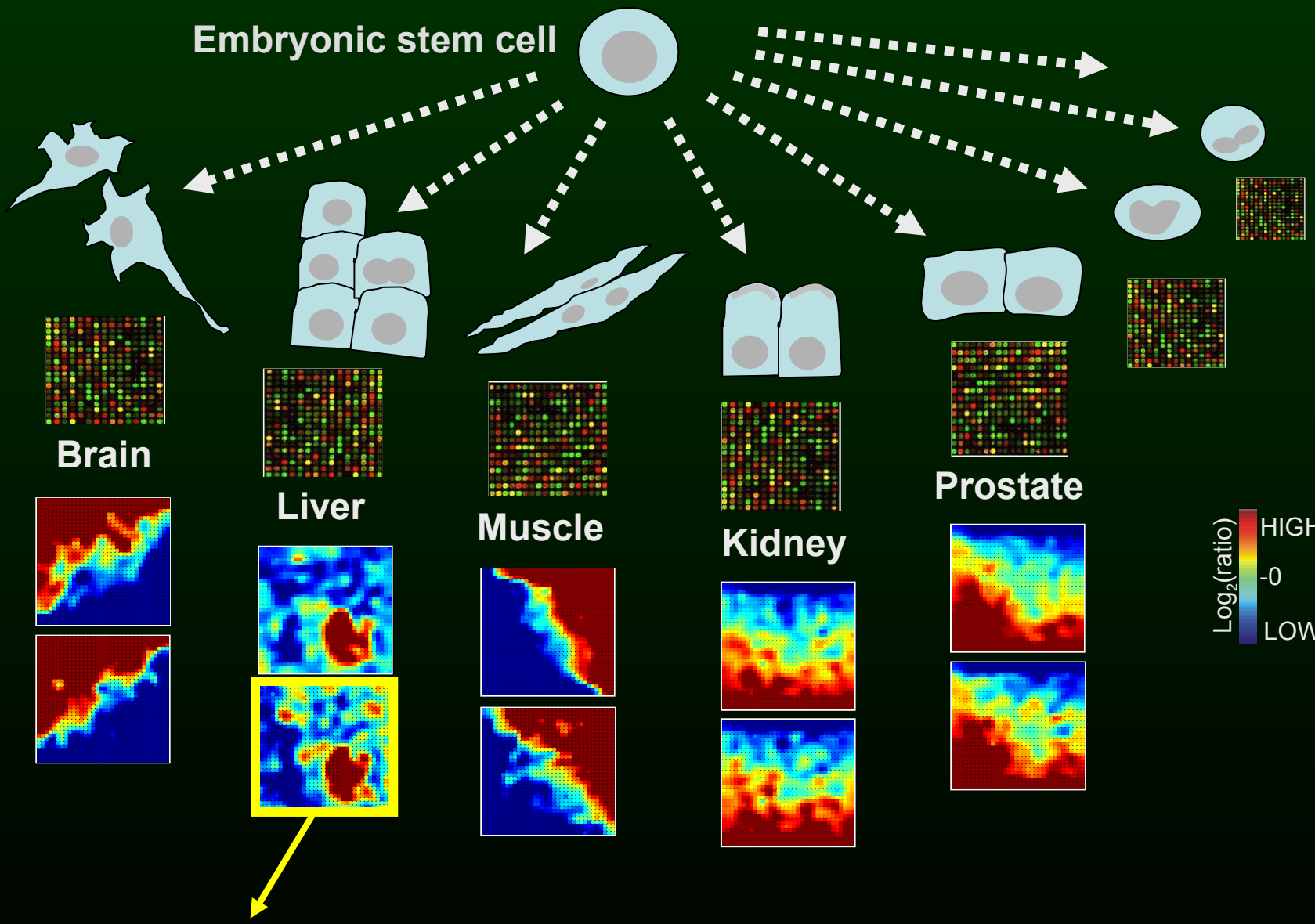
Spotted/Printed cDNA array
(on glass); 5000 genes



Affymetrix GeneChips®
(Photolithography); ~ 60,000 genes

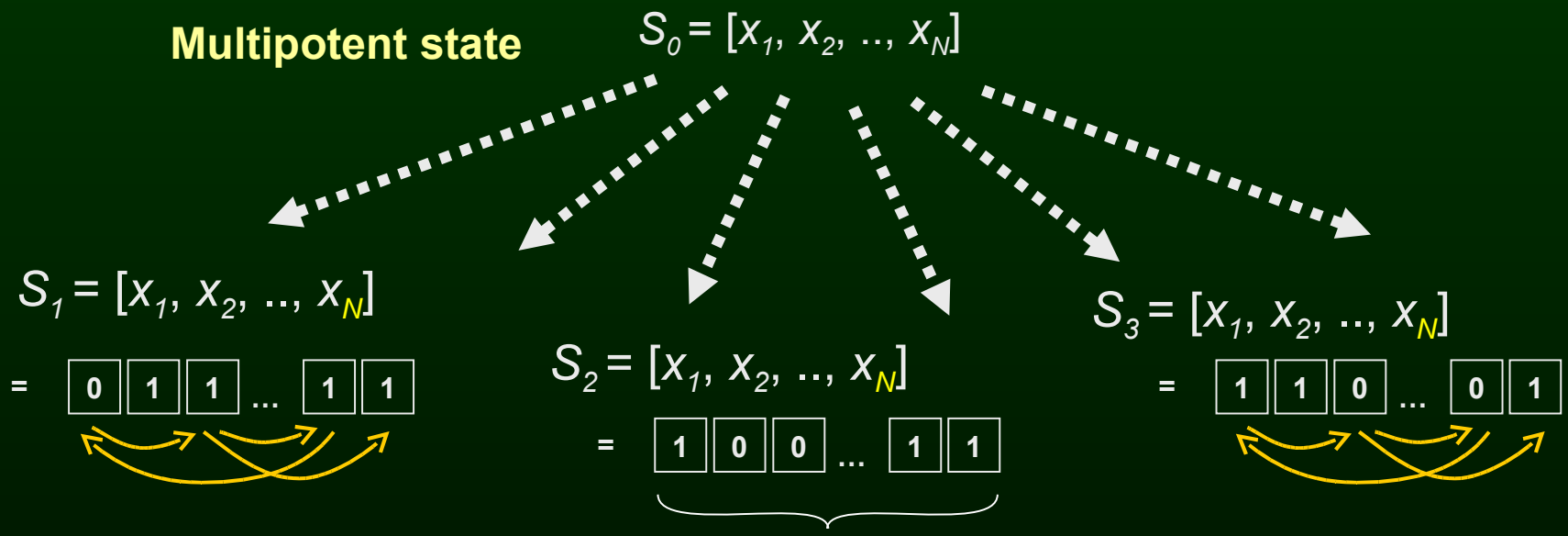
Formalizing the biological problem

The problem:



SOM-based "GEDI maps"
(Eichler et al., *Bioinformatics* 2003)

Formalizing the problem



If each gene is either ***ON*** or ***OFF*** (**1** or **0**)
and if there are N = 20,000 genes . . .

→ We would have $2^N = 2^{20,000} =$ **10^{6000}** gene expression constellations
(compare: there are 10^{80} hydrogen atoms in the universe)

→ **Yet, we observe only ~ 1000 discrete cell states / types**

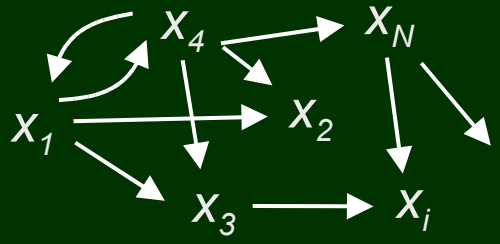
Network of regulatory interrelationships
constrains the possibilities of gene activation configurations

From Genes to System Behavior

One Genome

~ 100% "known"

One Network
(ARCHITECTURE)



~ 1-10% "known"

Whole Network Behavior
(DYNAMICS)

$$\dot{x}_1 = f_1(x_1, x_2 .. x_N)$$
$$\dot{x}_2 = f_2(x_1, x_2 .. x_N)$$
$$\vdots$$
$$\dot{x}_i = f_i(x_1, x_2 .. x_N)$$
$$\vdots$$
$$\dot{x}_N = f_N(x_1, x_2 .. x_N)$$

<< 1% "known"

"Biological Observable"
(PHENOTYPE)

Multiple stable states
Oscillations, Transitions

$$S(t_1) = [\, x_1(t_1),\, x_2(t_1),\, ..,\, x_N(t_1) \,]$$

$$\downarrow$$

$$S(t_2) = [\, x_1(t_2),\, x_2(t_2),\, ..,\, x_N(t_2) \,]$$

observable !

Biological observables to be explained

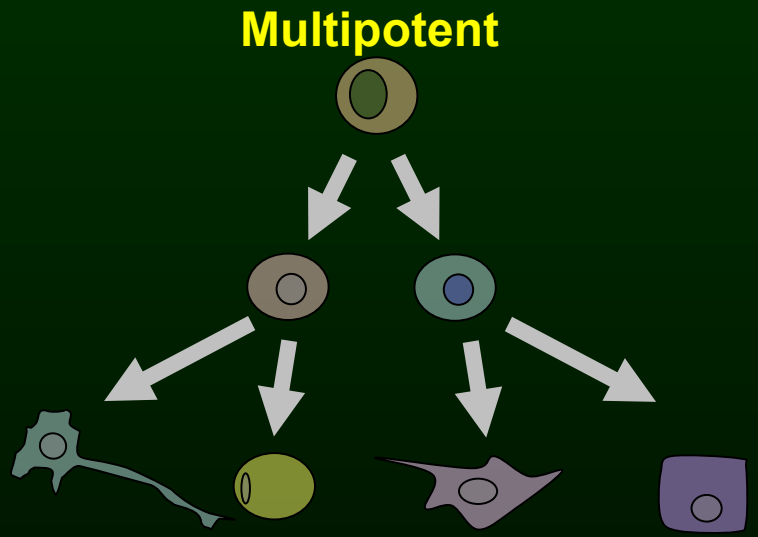
The four *D*'s:

DISCRETENESS

DIFFERENTIATION

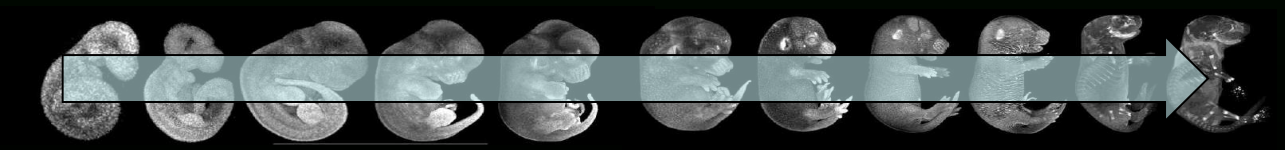
DIVERSIFICATION

DIRECTIONALITY



$S_0 = [x_1, x_2, .., x_N]$

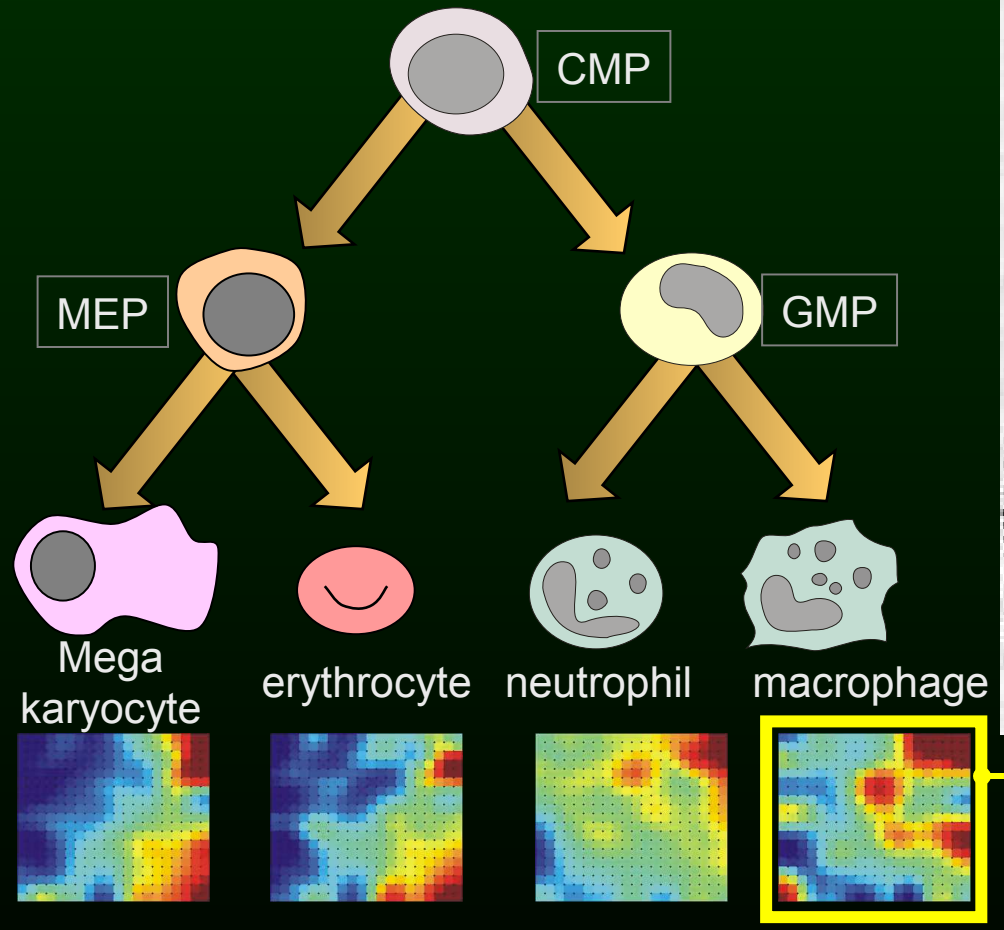
$S_1 = [x_1, x_2, .., x_N]$ $S_2 = [x_1, x_2, .., x_N]$

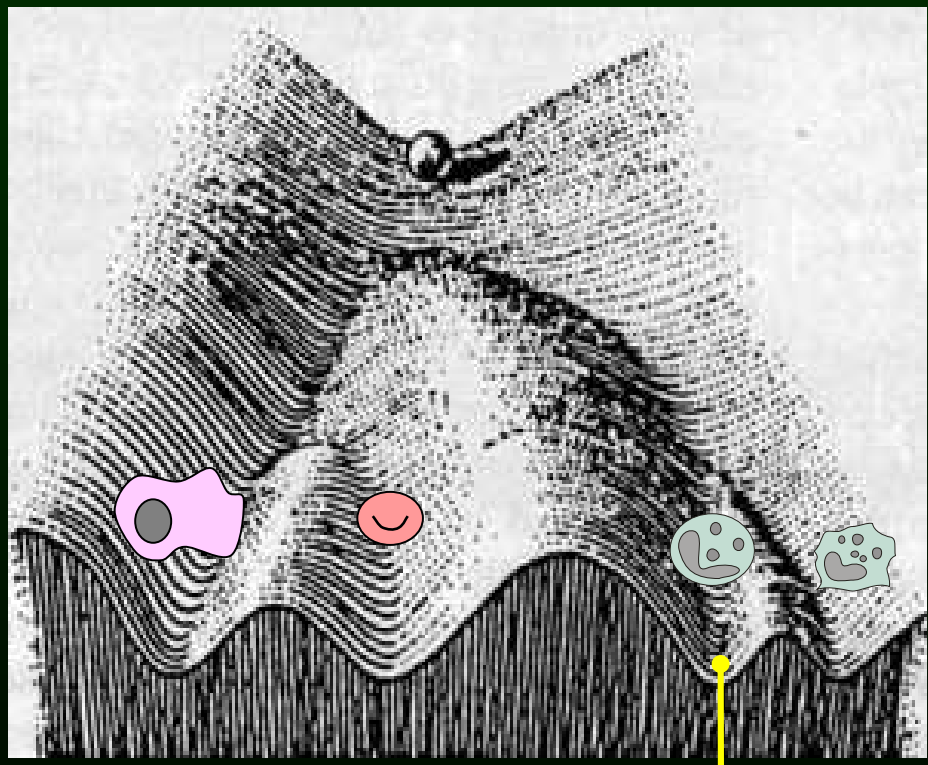


Epigenetic landscape and modern biology

The hematopoietic system

Multipotent progenitor cell





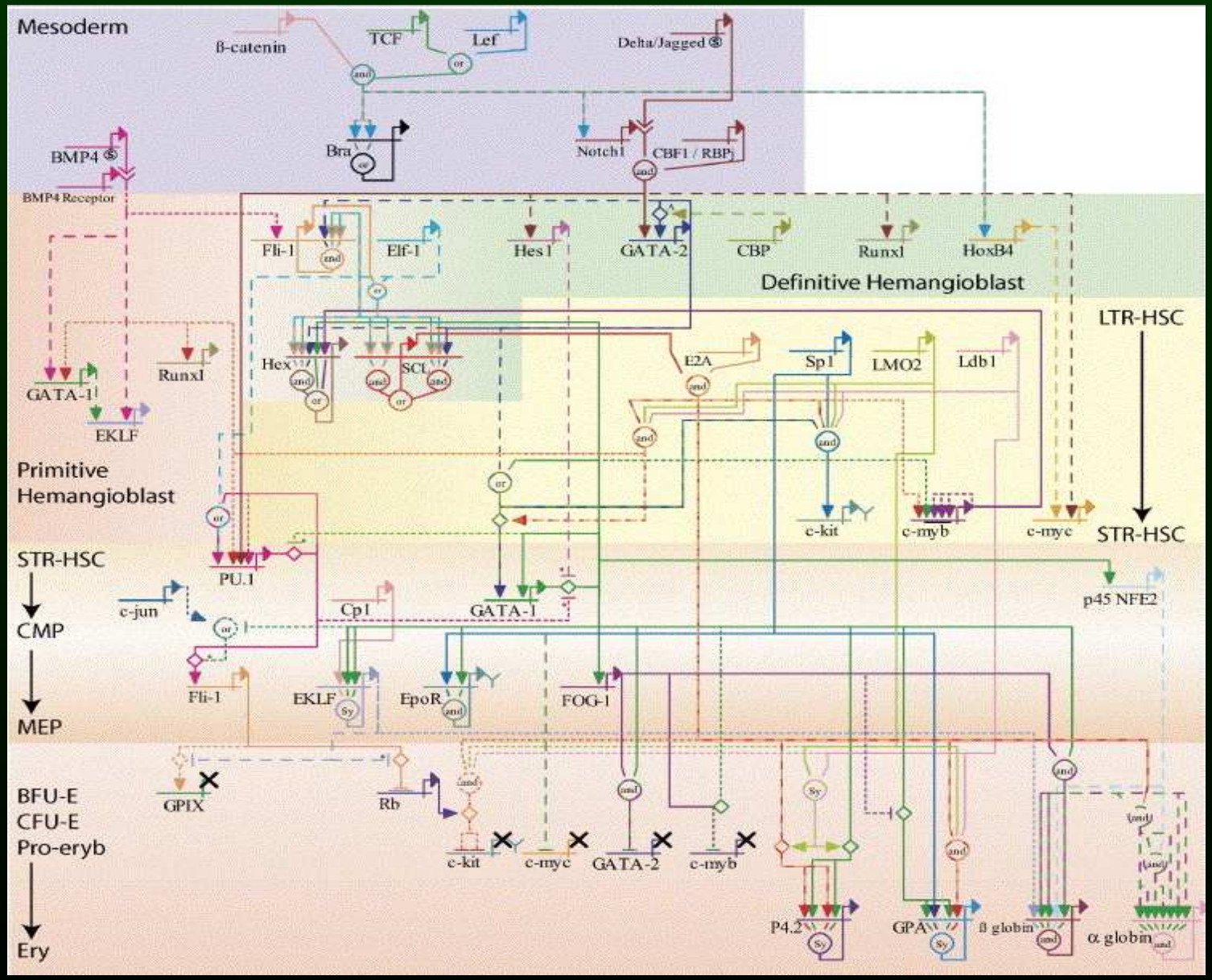
$S^*(t) = [\, x_1(t), x_2(t), \ldots x_N(t) \,]$

Stable state = cell type

More than a metaphor!

The *valleys* in the epigenetic landscape represent the stable gene expression profile which correspond to discrete *cell fates*

The gene regulatory network



From: Swiers et al. Dev. Biol. 2006 Control logic model of erythroid development

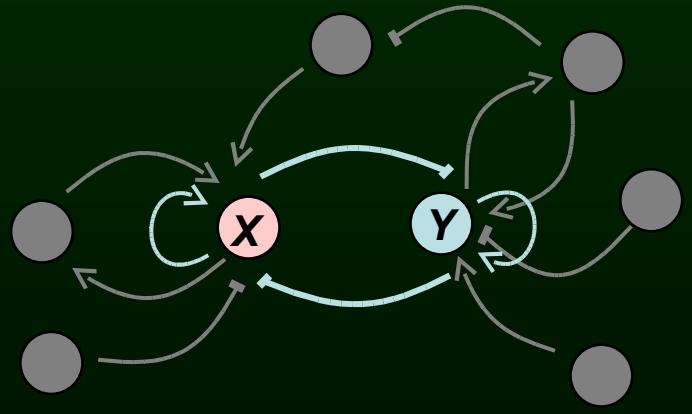
From network to landscape

Low vs. High-dimensional system

Dilemma:

Tractable: ⟷ Reality:

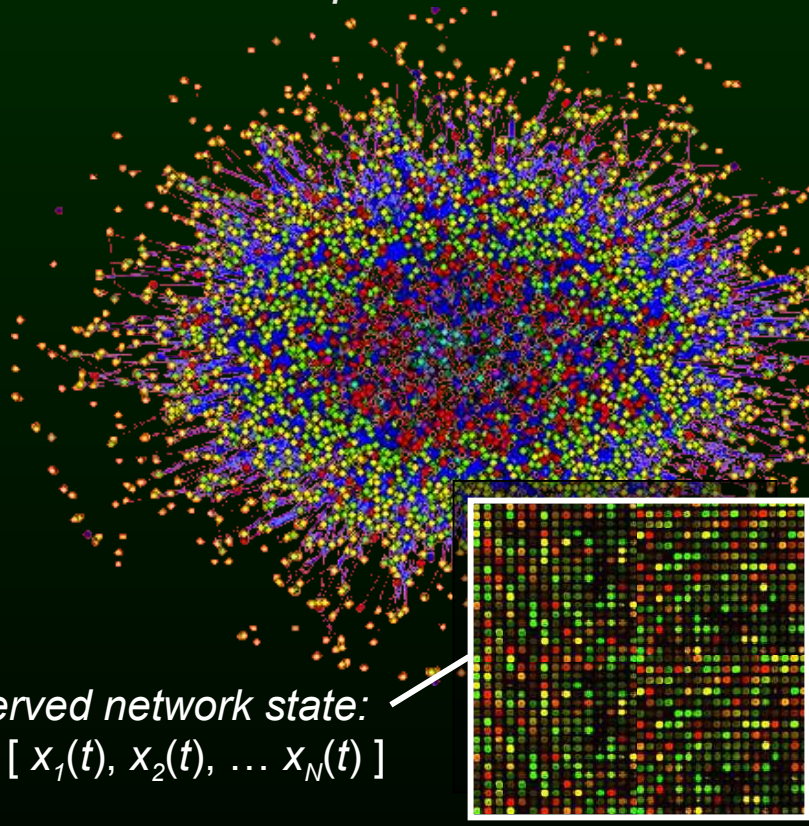
Small circuits



$$\frac{dX}{dt} = k_1 \frac{S_X{}^n}{S_X{}^n + Y^n} + k_2 \frac{X^n}{S_X{}^n + X^n} - k_3 X$$

$$\frac{dY}{dt} = k_3 \frac{S_Y{}^n}{S_Y{}^n + X^n} + k_5 \frac{Y^n}{S_Y{}^n + Y^n} - k_6 Y$$

Complex network

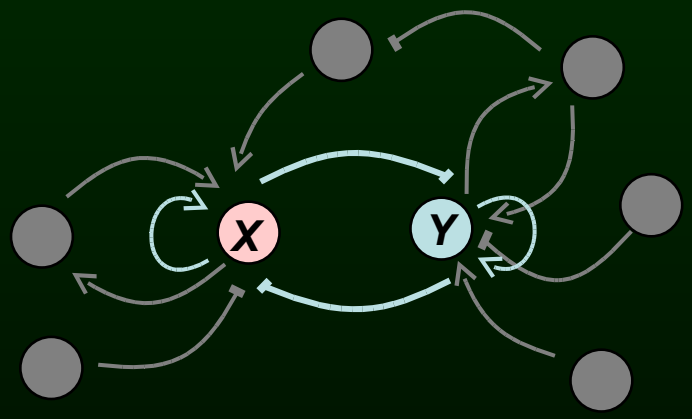


Observed network state:

$\boldsymbol{S}(t) = [\, x_1(t), x_2(t), \ldots x_N(t) \,]$

Local circuits are embedded in a genome-wide network

Low vs. **High** - dimensional system

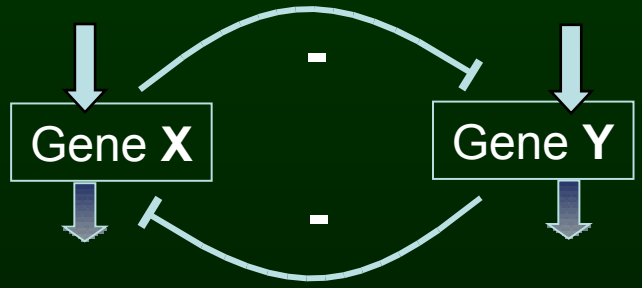


$$\frac{dX}{dt} = k_1 \frac{S_X{}^n}{S_X{}^n + Y^n} + k_2 \frac{X^n}{S_X{}^n + X^n} - k_3 X$$

$$\frac{dY}{dt} = k_3 \frac{S_Y{}^n}{S_Y{}^n + X^n} + k_5 \frac{Y^n}{S_Y{}^n + Y^n} - k_6 Y$$

Let’s start small . . .

BASICS: Two–gene network - mutual inhibition



$$\frac{dX}{dt} = k_1 \frac{S_X^{\,n}}{S_X^{\,n} + Y^n} - k_2 X \quad \text{+ noise}$$

$$\frac{dY}{dt} = k_3 \frac{S_Y^{\,n}}{S_Y^{\,n} + X^n} - k_4 Y \quad \text{+ noise}$$

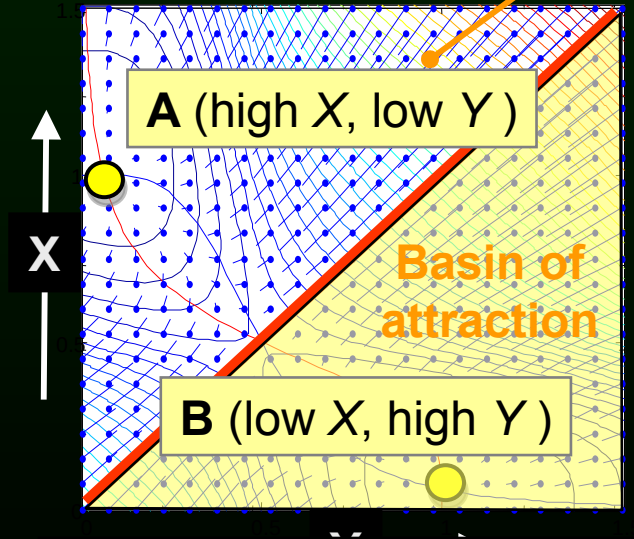
The global dynamics of this system:

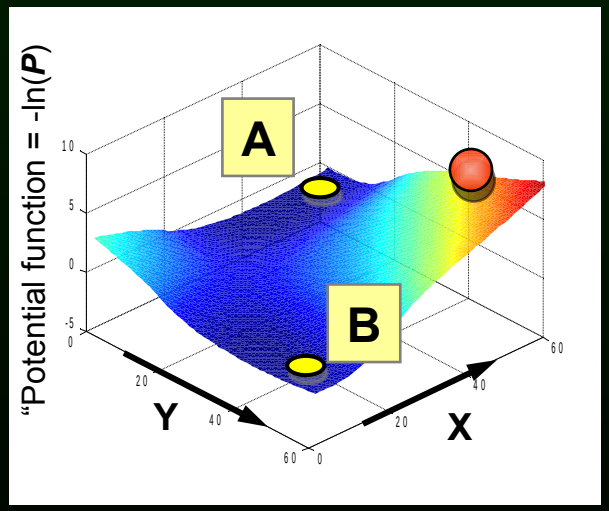
$S(t) = [\, X(t), Y(t) \,]$

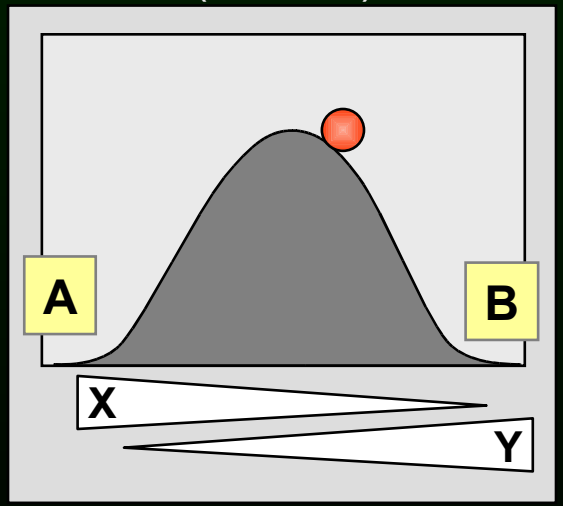
2-D **state space**

as "potential landscape"

(section)

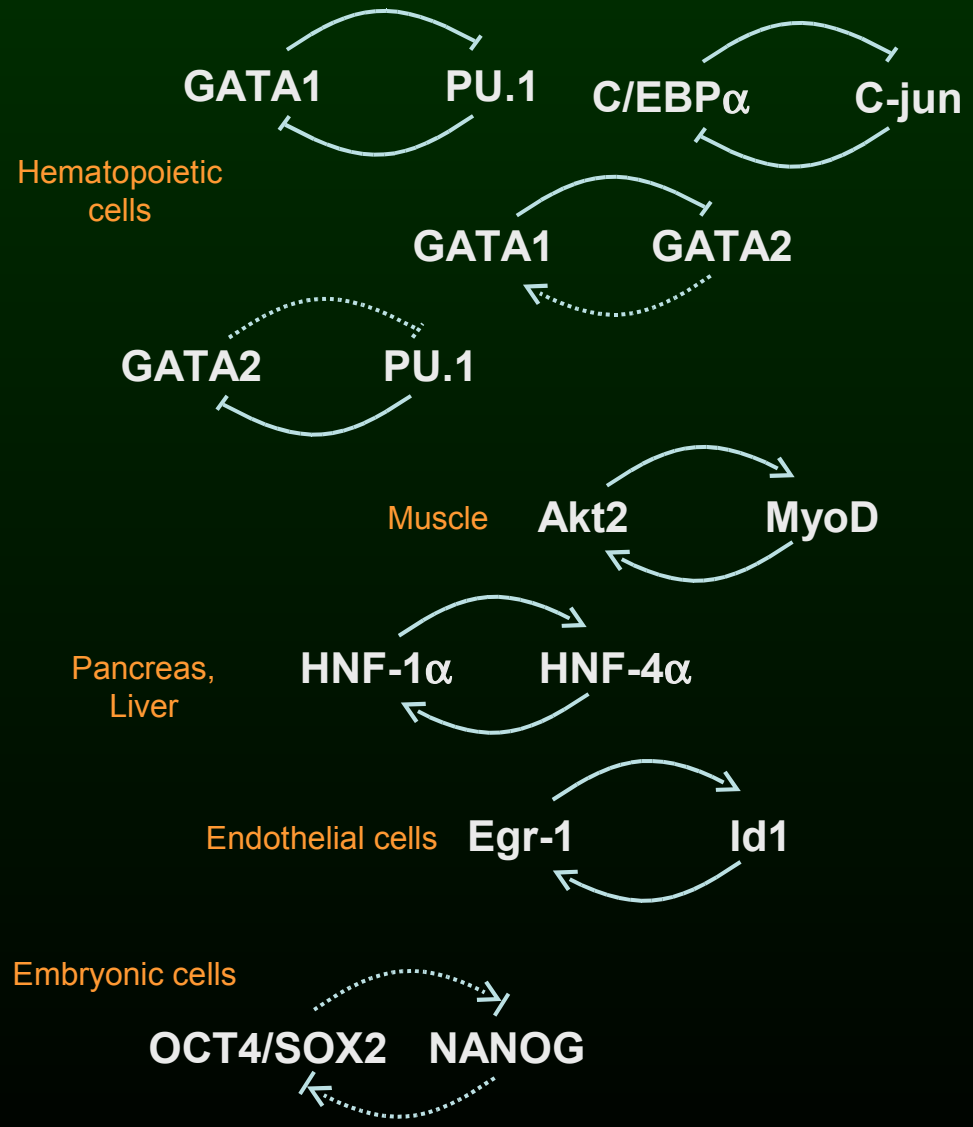




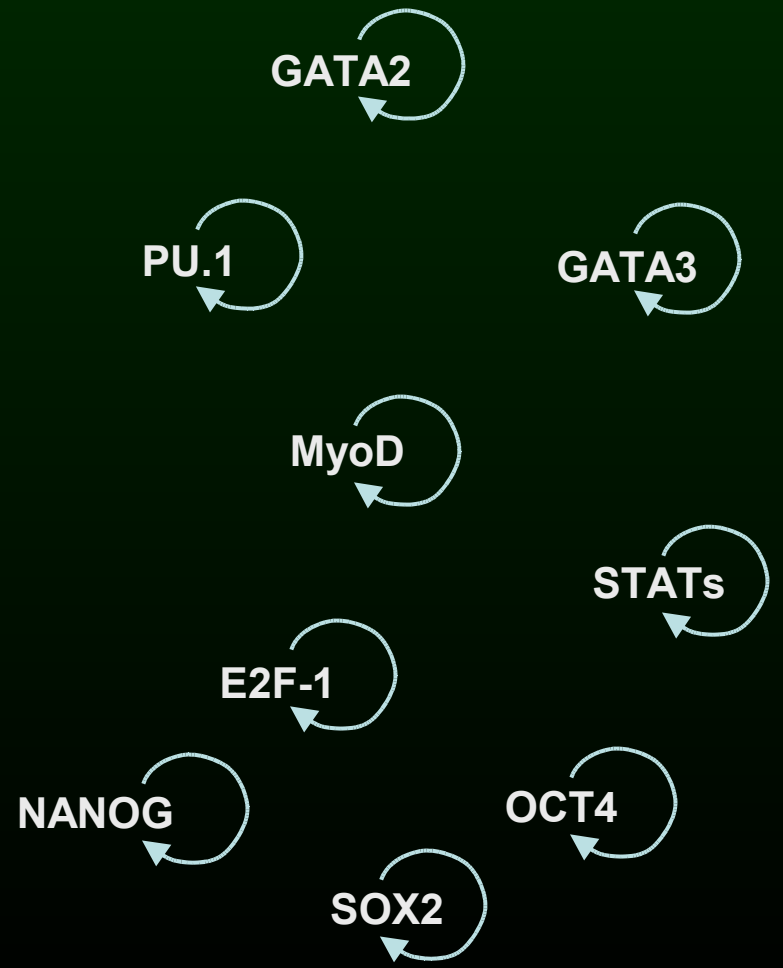


= Bistability (Multistability).

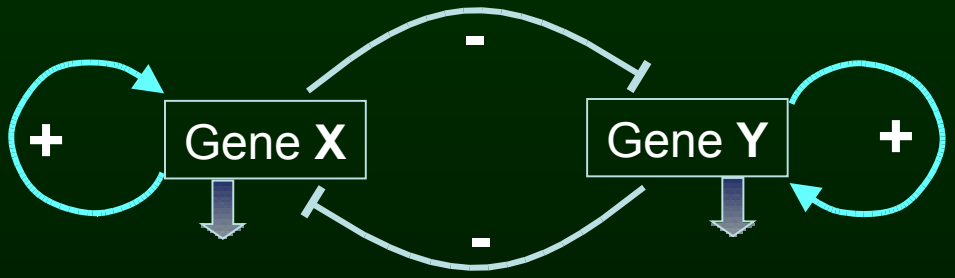
Feedback “modules” in transcriptional regulation



Auto-regulation:
Transcription factors
That bind their own promoters

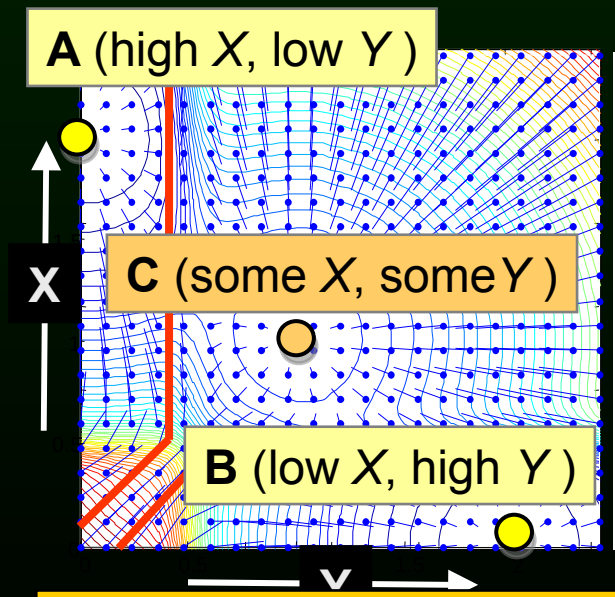


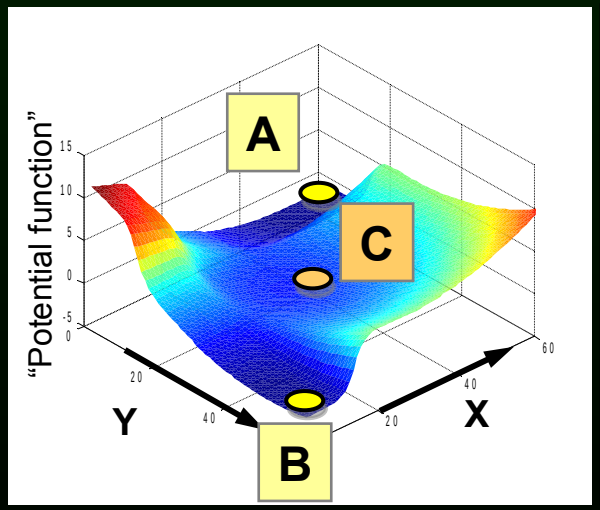
Two–gene network: mutual inhibition + auto-stimulation

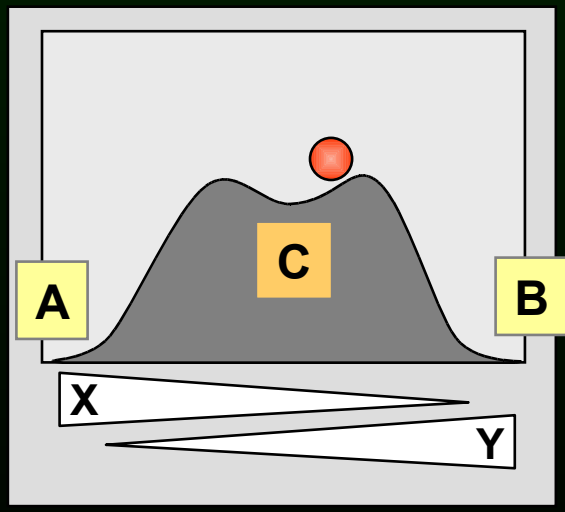


$$\frac{dX}{dt} = k_1 \frac{S_X{}^n}{S_X{}^n + Y^n} + k_2 \frac{X^n}{S_X{}^n + X^n} - k_3 X$$

$$\frac{dY}{dt} = k_3 \frac{S_Y{}^n}{S_Y{}^n + X^n} + k_5 \frac{Y^n}{S_Y{}^n + Y^n} - k_6 Y$$







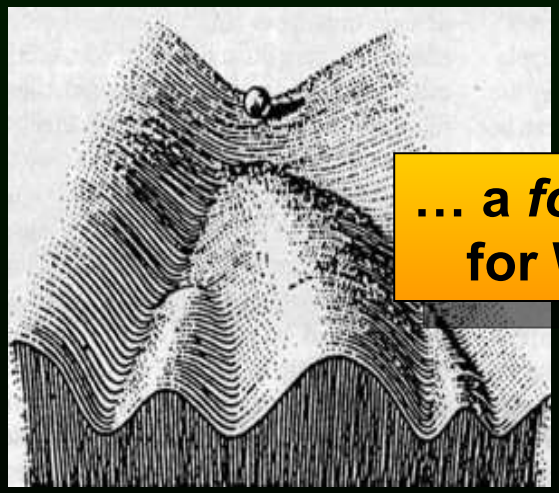
Huang, Comp. Systems Biol. (2005)

= Tri-stability

THE CENTRAL HYPOTHESIS :

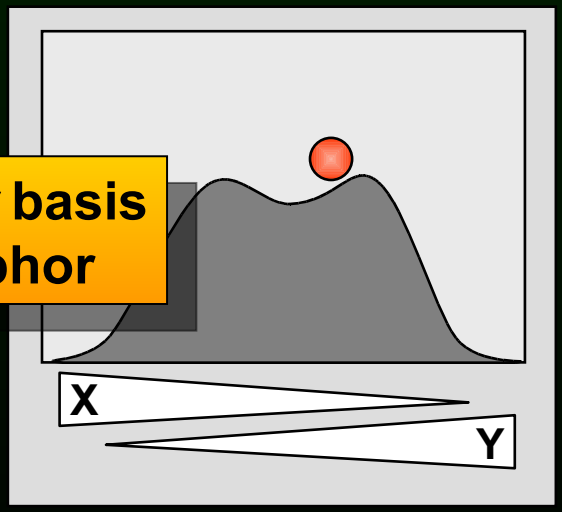
" CELL TYPES ARE ATTRACTORS OF THE GENE REGULATORY NETWORK "

- 1949 **Delbrück**
- 1961 **Jacob and Monod**
- 1969 **Stuart Kauffman** (complex networks)



1940s Waddington:
" Epigenetic landscape "

… a *formal* and *molecular* basis for Waddington's metaphor



Multistability

From small circuits to complex networks

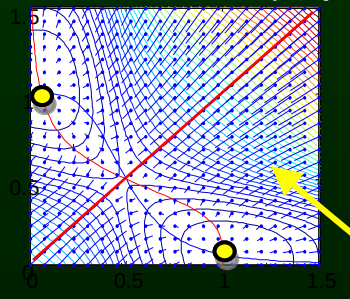
TOPOLOGY
(“ wiring diagram “)

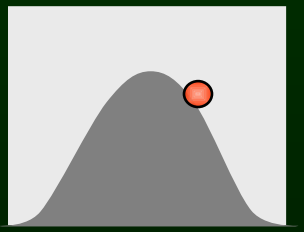
DYNAMICS
(“epigenetic landscape”)

$N = 2$

$$\frac{dX}{dt} = k_1 \frac{S_X^{\,n}}{S_X^{\,n} + Y^n} - k_2 X$$

$$\frac{dY}{dt} = k_3 \frac{S_Y^{\,n}}{S_Y^{\,n} + X^n} - k_4 Y$$

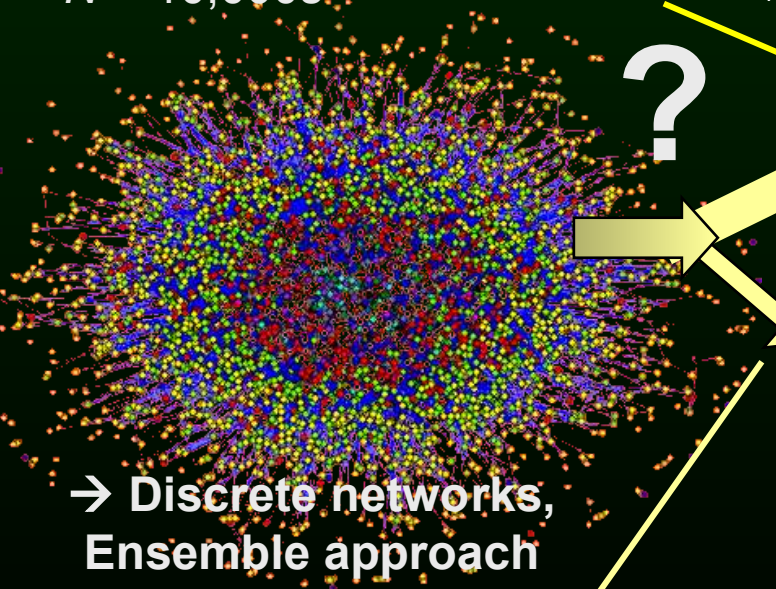




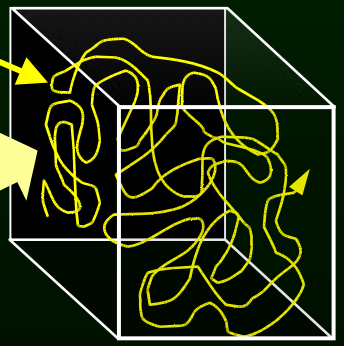
$\boldsymbol{S}(t) = [\, X(t),\ Y(t) \,]$

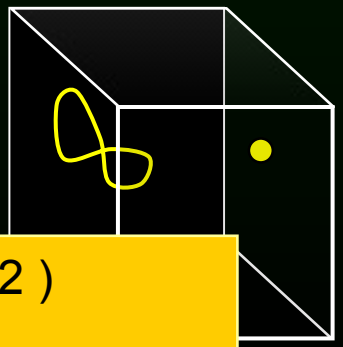
N = 10,000s

$\boldsymbol{S}(t) = [\, x_1(t),\ x_2(t),\ \ldots\ x_N(t) \,]$



?





“CHAOTIC” (Derrida, Kauffman)
~UNSTABLE
Very long transients or
Very high-number of tiny attractors

CRITICAL

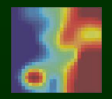
ORDERED
~ STABLE / DISSIPATIVE
Relatively small number
of compact attractors
(fixed-point or short-period cycles)

→ Discrete networks, Ensemble approach

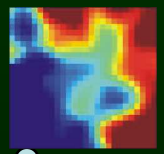
- Sparse networks ($k = 2$)
- “scale-free” networks
- Canalizing functions (→ miRNA!)

(Kauffman, Aldana, Shmulevich et al.)

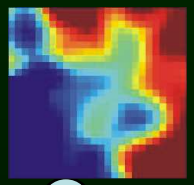
N-dimensional hyper space

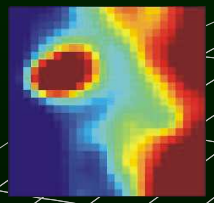


THEORY:
With 25,000 genes (ON/OFF):
→ $2^{25,000} \sim 10^{7500}$ expression profiles:
→ a huge continuum of configurations

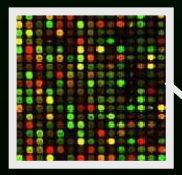


BIOLOGICAL REALITY: Only a tiny fraction of possible states are stable and realized



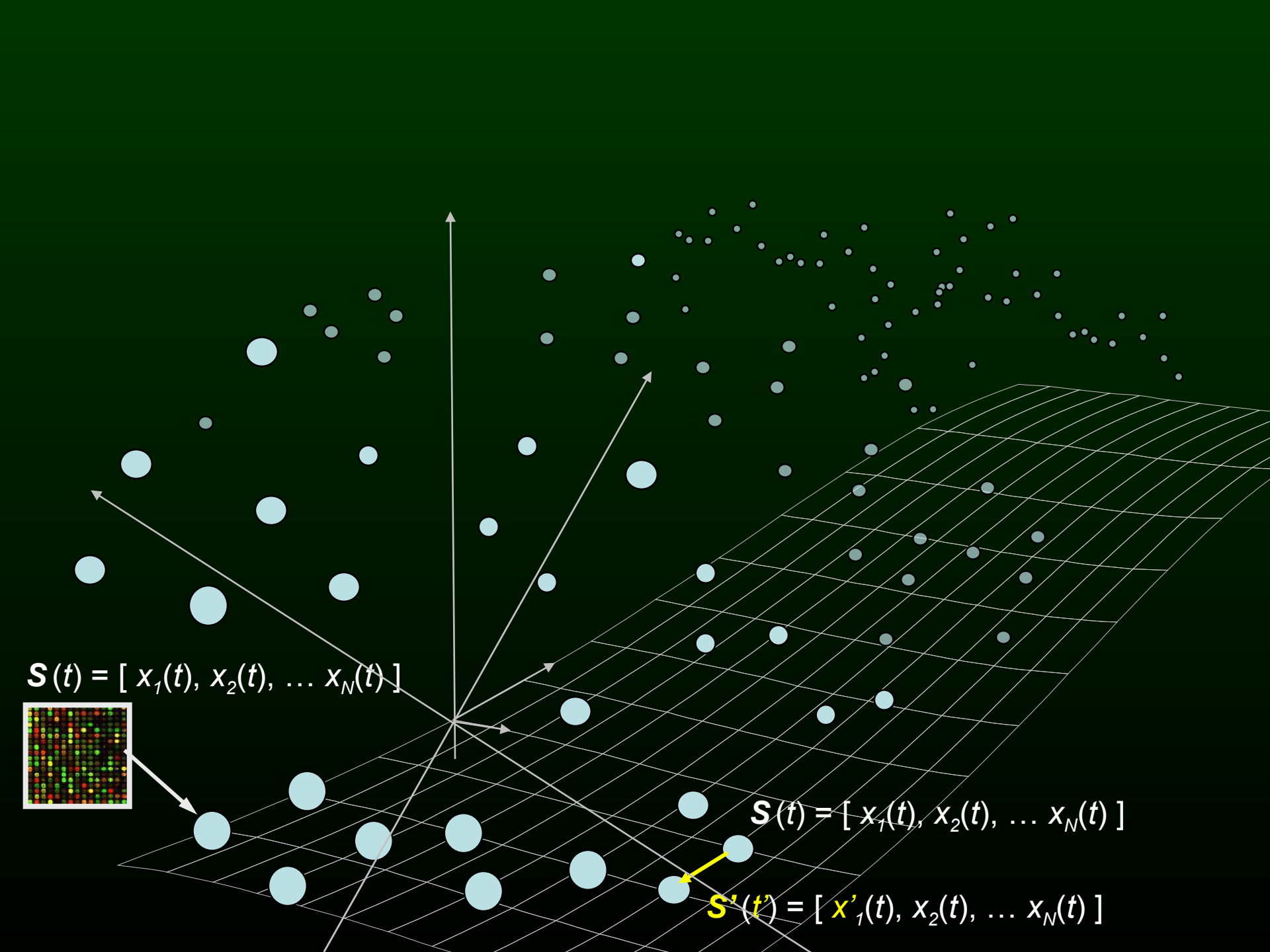


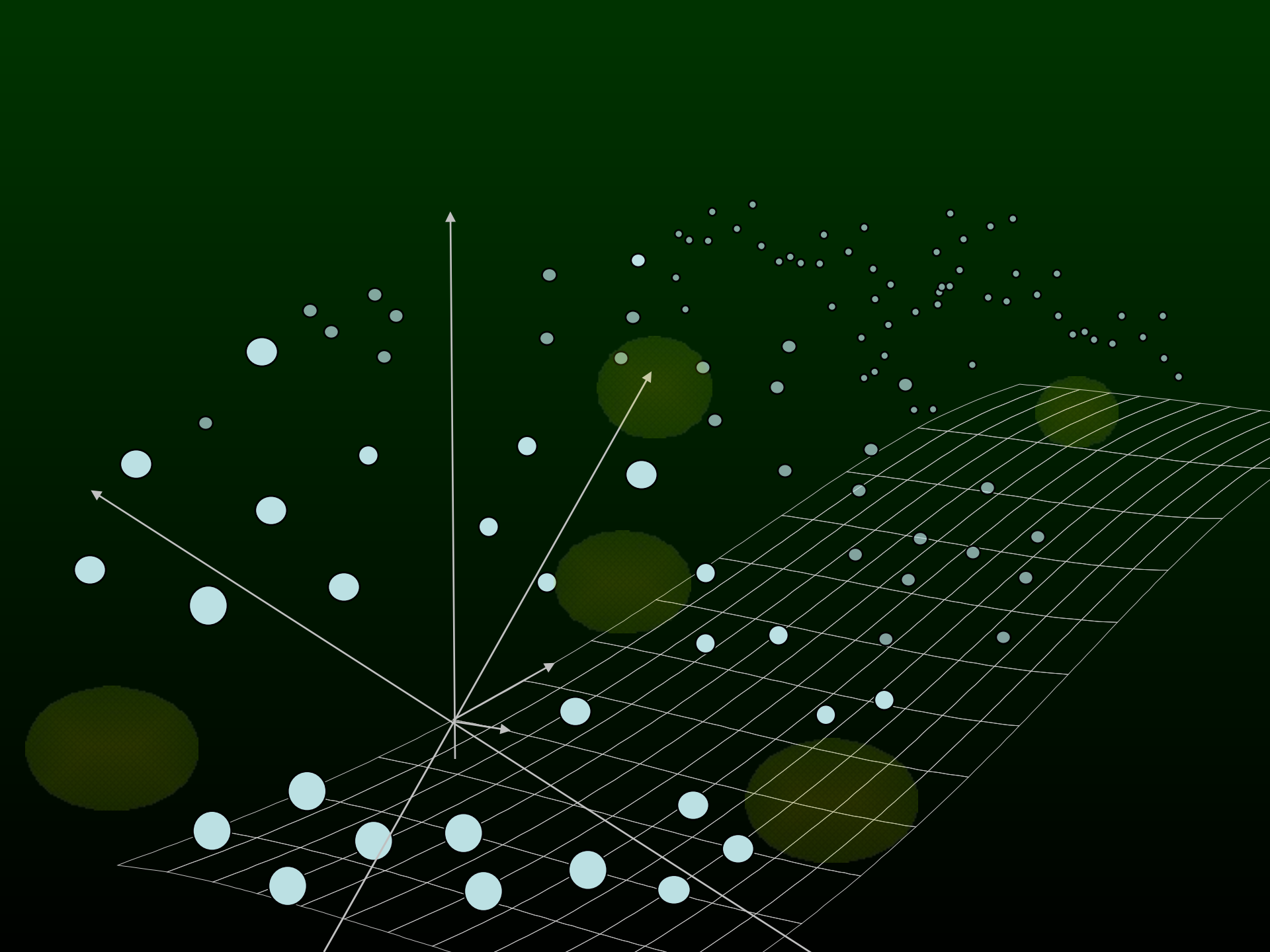
$S(t) = [\, x_1(t), x_2(t), \ldots x_N(t) \,]$



$S(t) = [\, x_1(t), x_2(t), \ldots x_N(t) \,]$

$S'(t') = [\, x'_1(t), x_2(t), \ldots x_N(t) \,]$


S (t) = [x1(t), x2(t), … xN(t)]
S (t) = [x1(t), x2(t), … xN(t)]
S' (t') = [x'1(t), x2(t), … xN(t)]

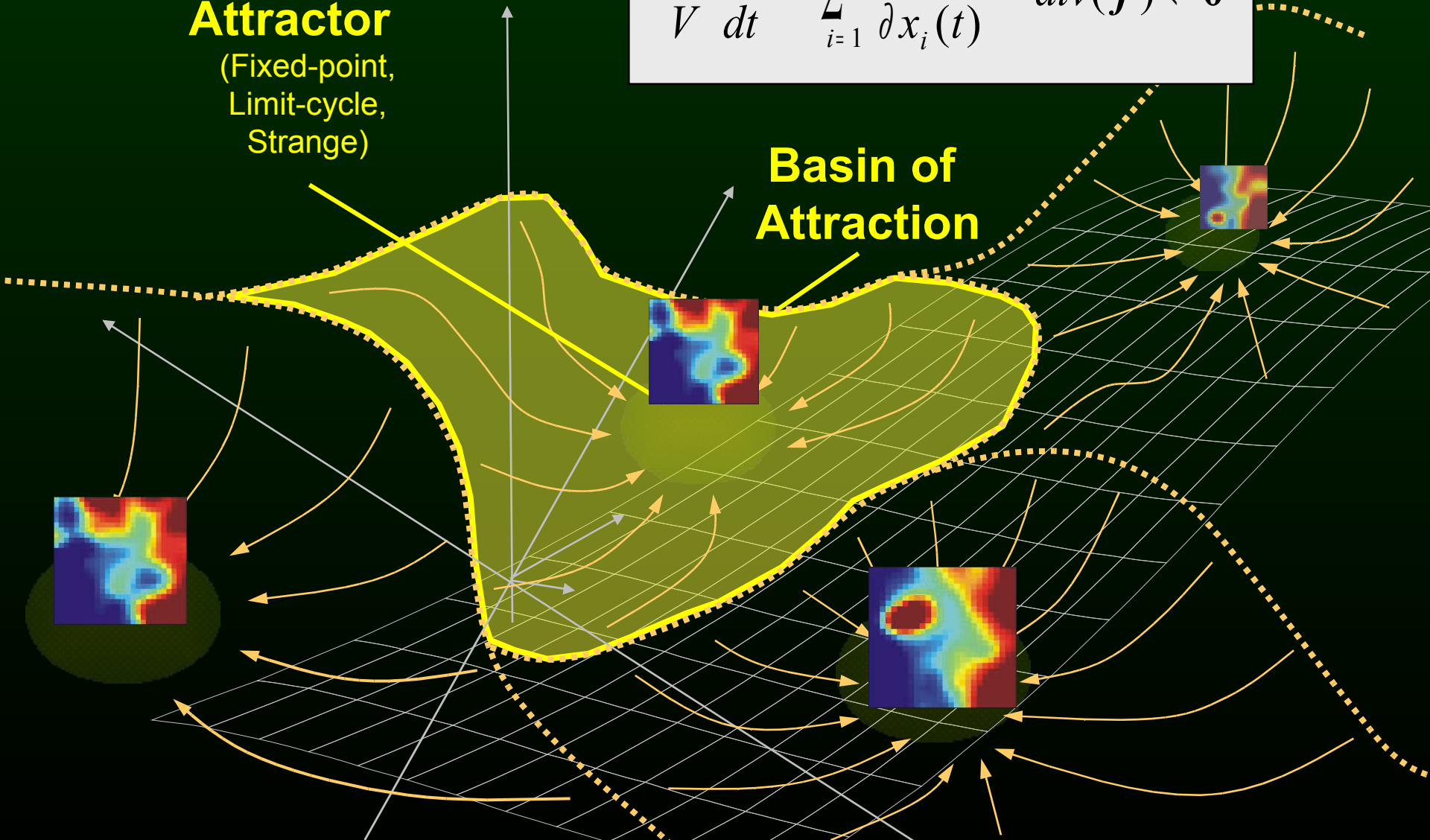


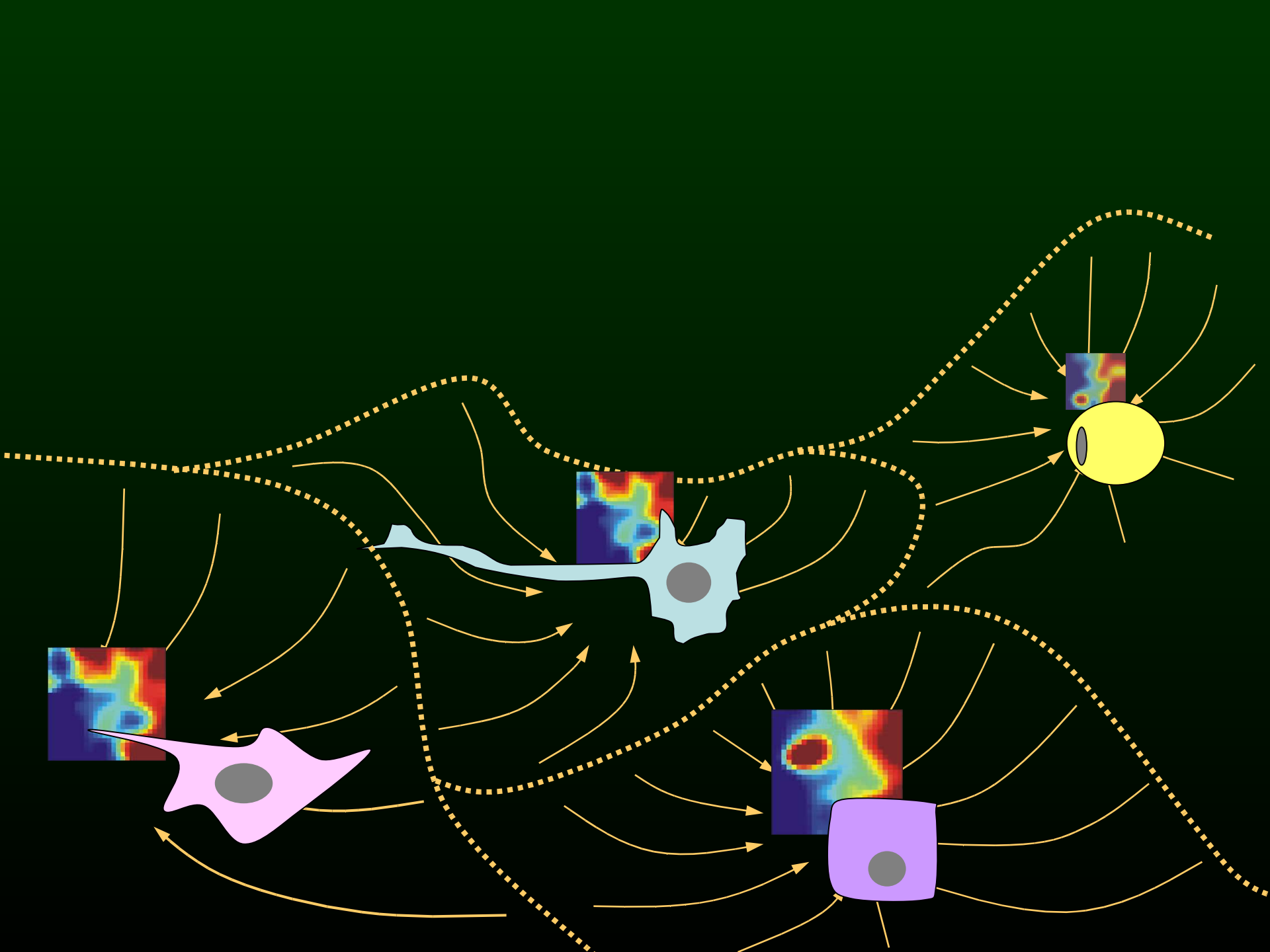
Dissipative system:

$$\frac{1}{V}\frac{dV}{dt} = \sum_{i=1}^{N} \frac{\partial f_i(x)}{\partial x_i(t)} = div(\boldsymbol{f}) < \mathbf{0}$$

Attractor
(Fixed-point,
Limit-cycle,
Strange)

Basin of Attraction





Biological observables to be explained

The four *D*'s:

DISCRETENESS 

DIFFERENTIATION

DIVERSIFICATION

DIRECTIONALITY

Biological observables to be explained

The four *D*'s:

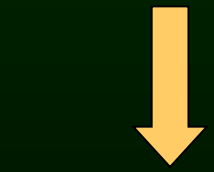
DISCRETENESS ✓

DIFFERENTIATION

DIVERSIFICATION

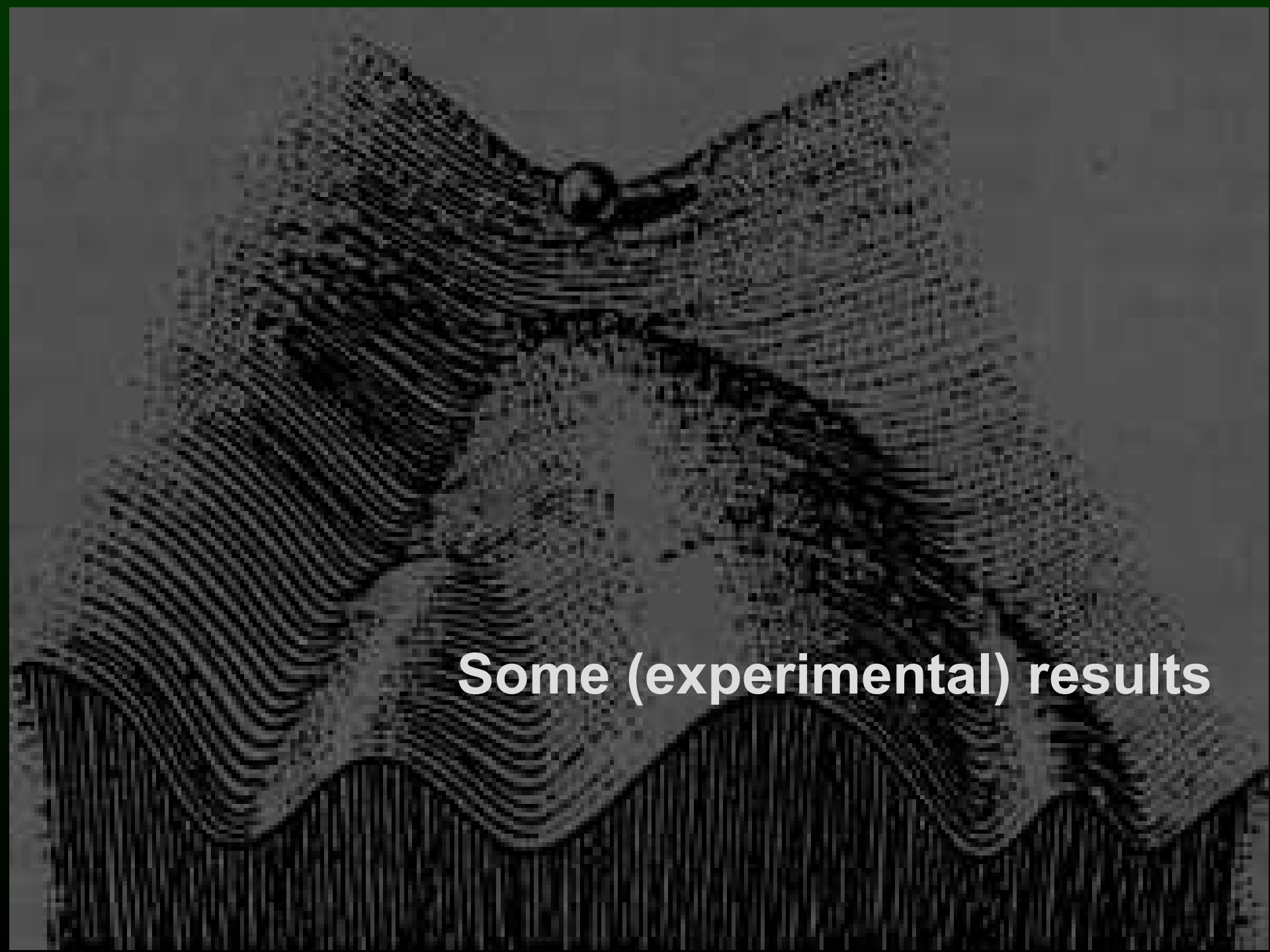
DIRECTIONALITY

$S_1 = [x_1, x_2, .., x_N]$

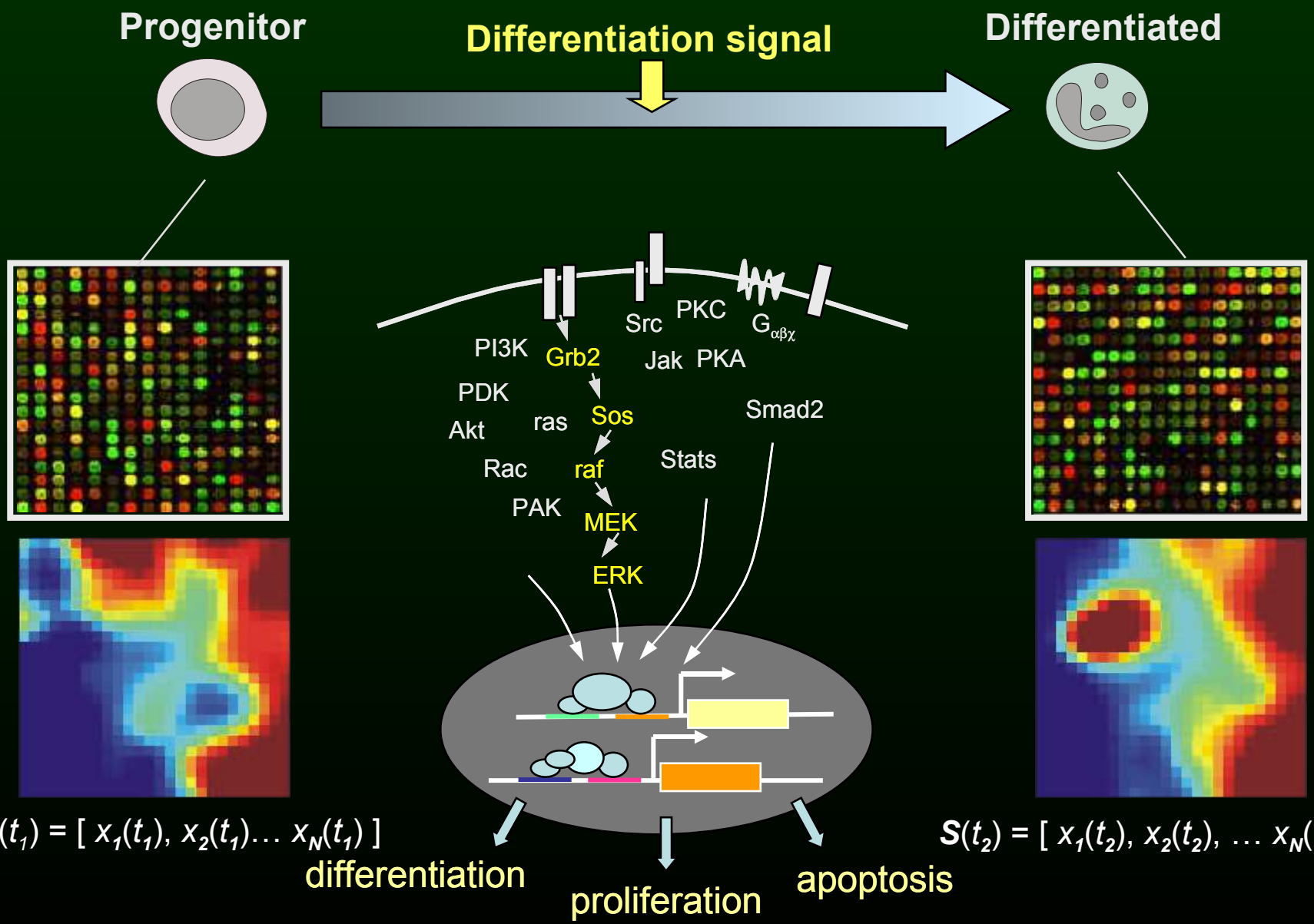


$S_2 = [x_1, x_2, .., x_N]$

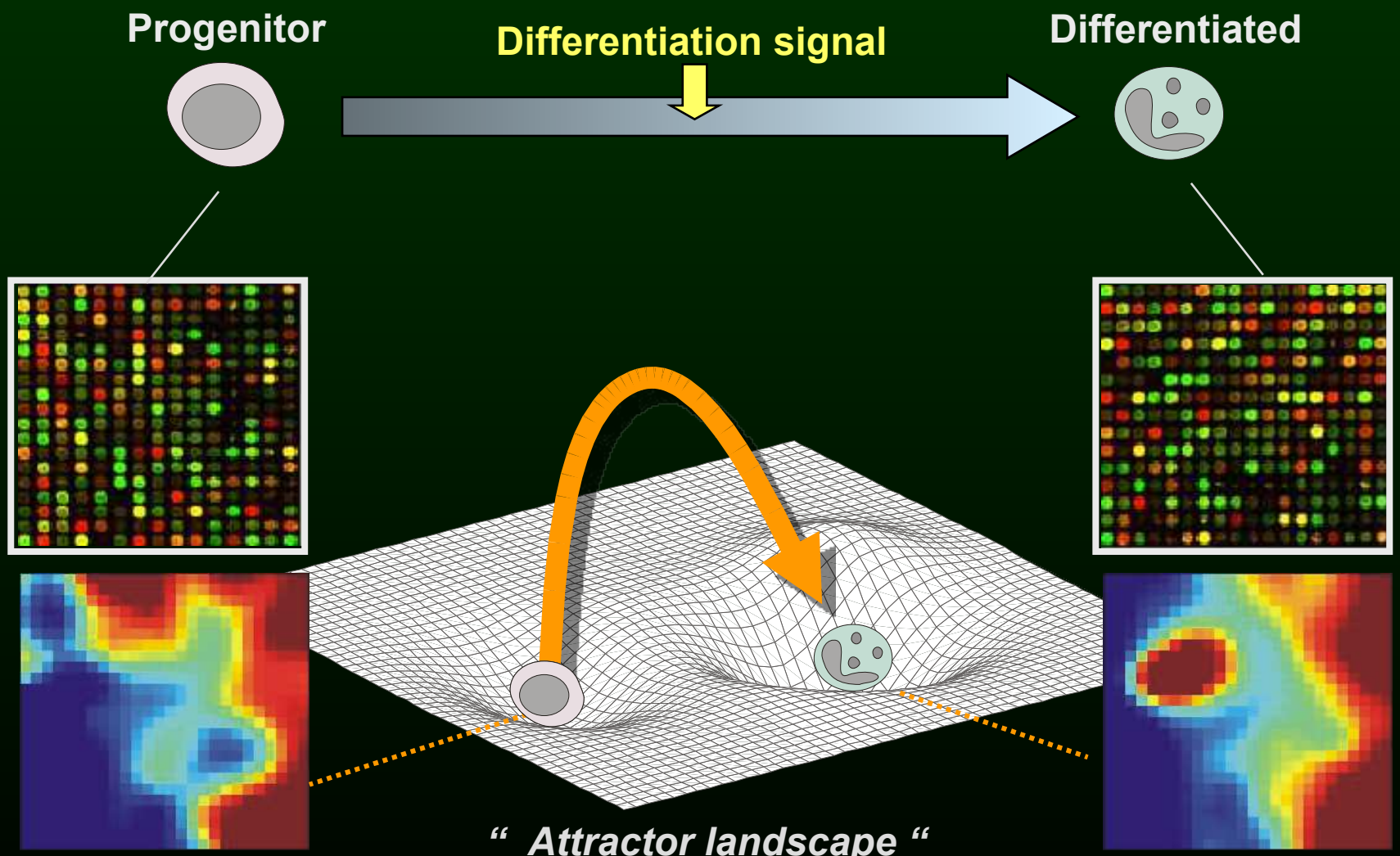
Is a differentiation a switch into a high-dimensional *attractor* state?


Some (experimental) results

Differentiation as a state transition



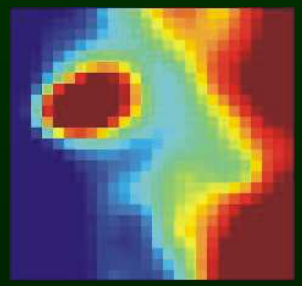
GEDI - Eichler et al., Bioinformatics (2003)



$S^{Prog}(t) = [\, x_1{}^{Prog}(t), \ldots x^{Prog}{}_N(t) \,]$

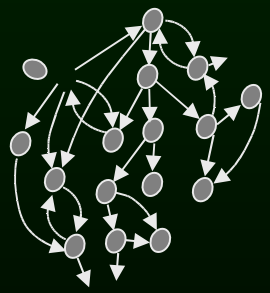
$S^{Diff}(t) = [\, x_1{}^{Diff}(t), \ldots x^{Diff}{}_N(t) \,]$

When is a *stationary* state a *stable* (attractor) state?



Is a measured, ***stationary*** gene expression profile
$\boldsymbol{S}^*(t) = [x_1(t), x_2(t), \ldots x_N(t)]$
a "***stable***" state ?

If network architecture known $\dot{\boldsymbol{x}} = \boldsymbol{f}(\boldsymbol{x})$



1. Jacobian matrix → Eigenvalues: all <0
   (for small neighborhood of S*)
2. Lyapunov function $V(S)$:

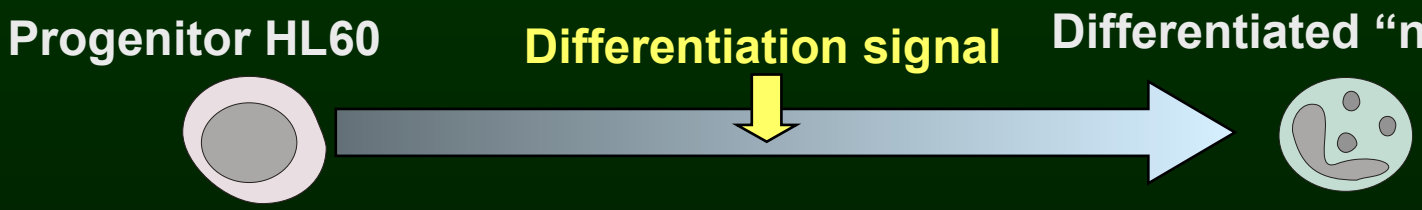
$V(S) > 0$ for $S \neq S^*$
$V(S^*) = 0$
$\dot{V}(S) = \nabla V(S) \cdot \boldsymbol{f}(x) < 0$

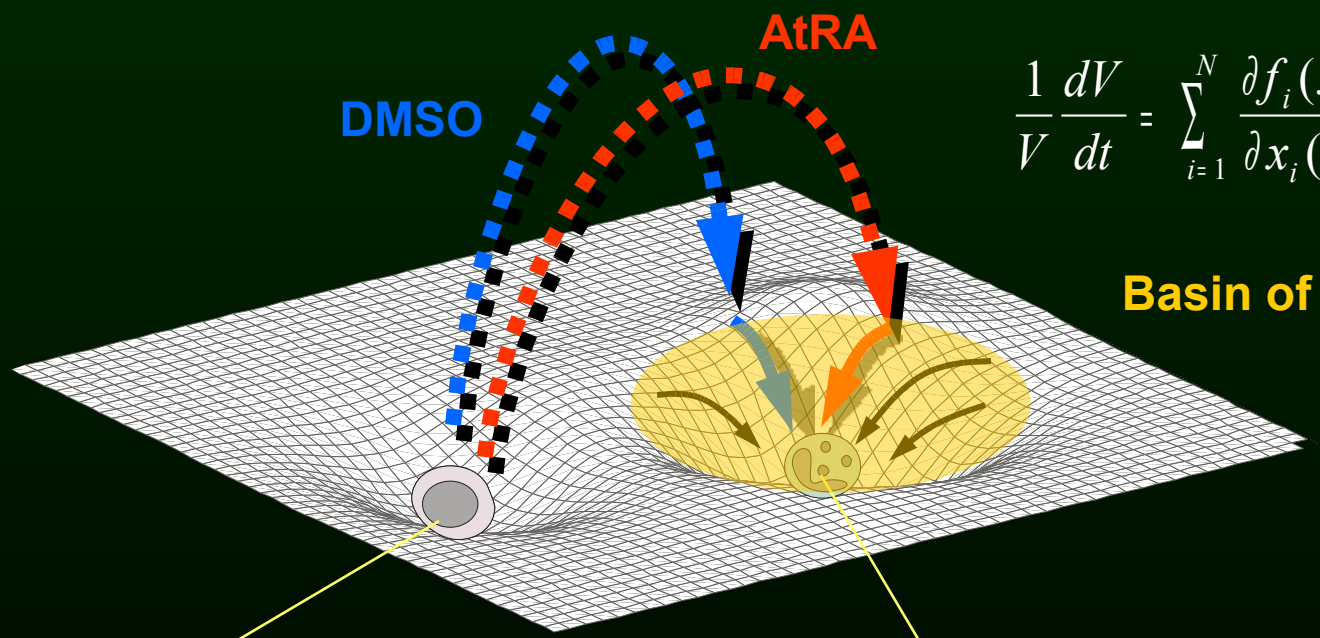
If network architecture unknown:

Realistically:
→ **Look for convergence of trajectories**

$$\frac{1}{V}\frac{dV}{dt} = \sum_{i=1}^{N} \frac{\partial f_i(x)}{\partial x_i(t)} = div(\boldsymbol{f}) < 0$$

What to expect from time course of expression profiles

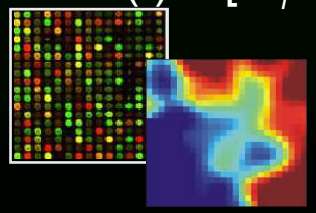




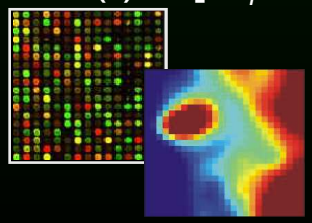
$$\frac{1}{V}\frac{dV}{dt} = \sum_{i=1}^{N} \frac{\partial f_i(x)}{\partial x_i(t)} = div(\boldsymbol{f}) < \mathbf{0}$$

Basin of attraction !

$S^{Prog}(t) = [\, x_1{}^{Prog}(t), \ldots x^{Prog}{}_N(t) \,]$



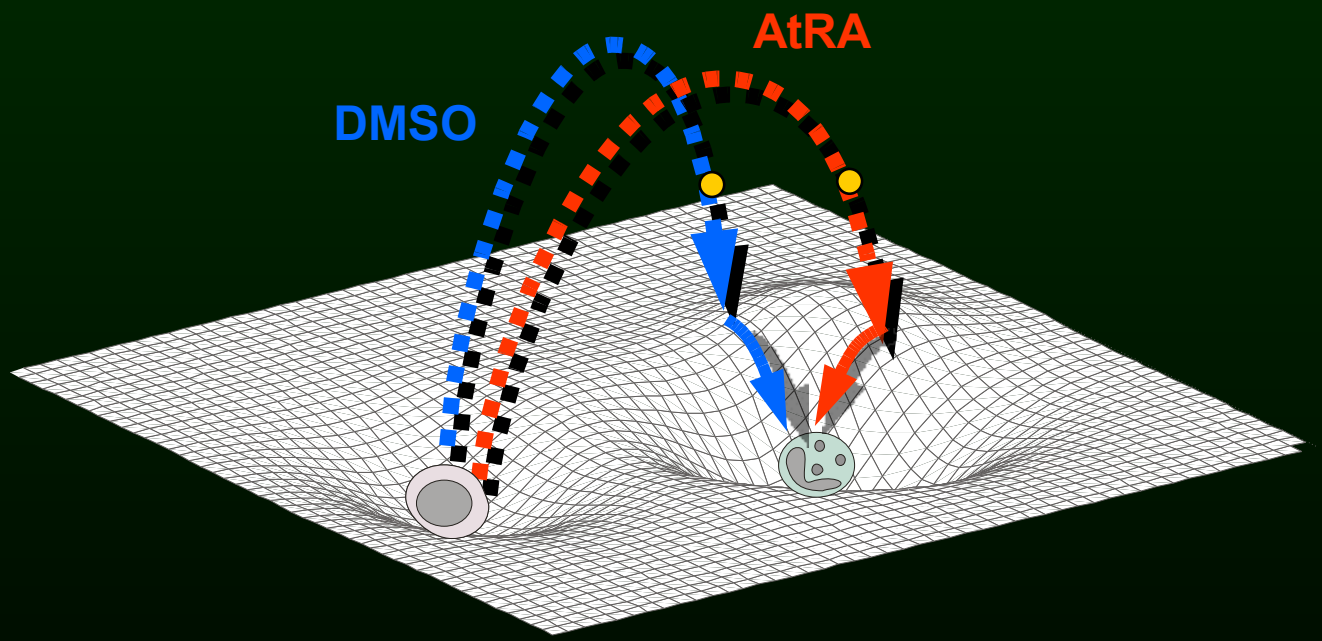
$S^{Diff}(t) = [\, x_1{}^{Diff}(t), \ldots x^{Diff}{}_N(t) \,]$



Progenitor HL60

Differentiation signal

Differentiated “neutrophil”

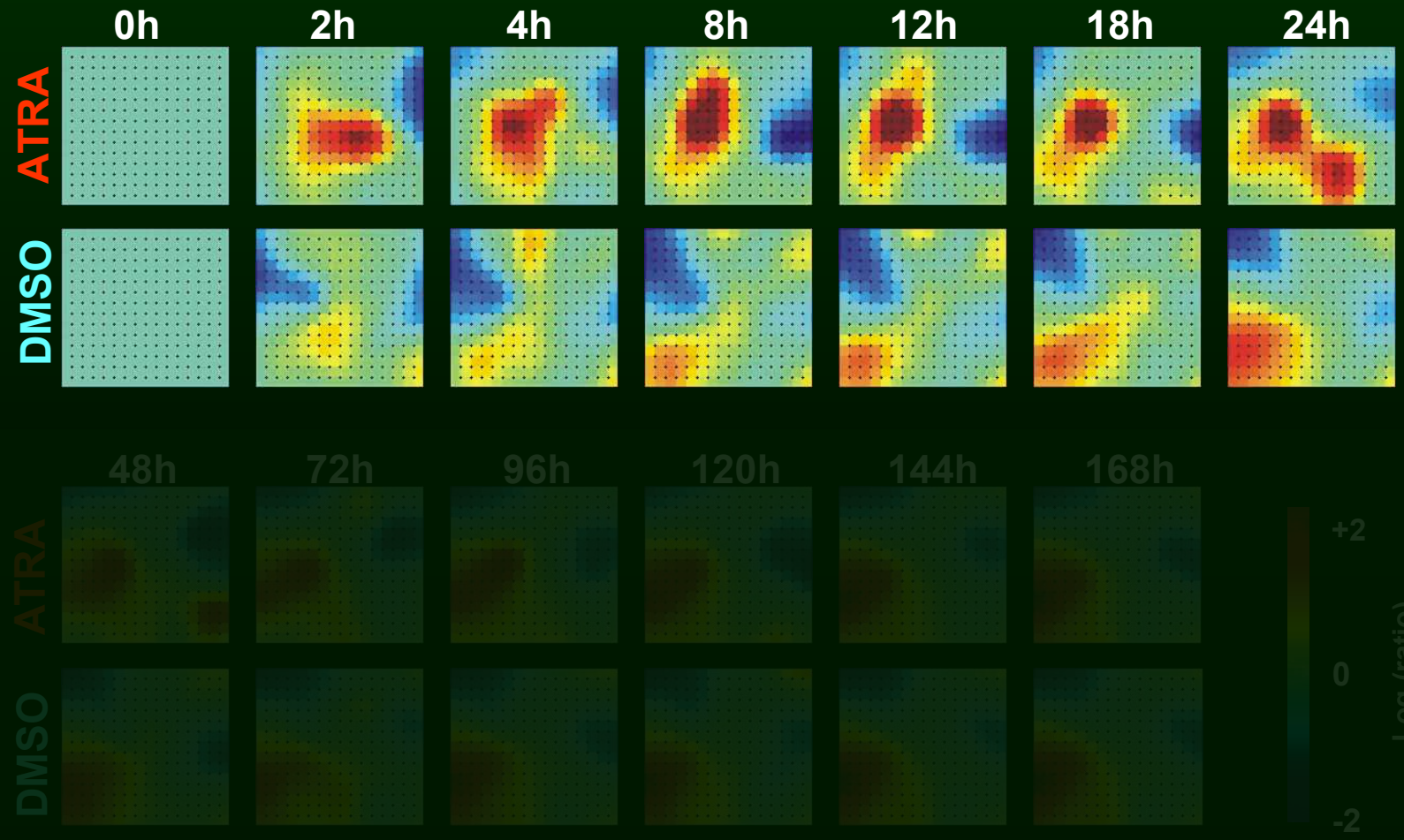


Look for:

- Initial divergence of trajectories, followed by ***convergence*** of trajectories
- Distributed contribution by a ***genome-scale*** number of genes

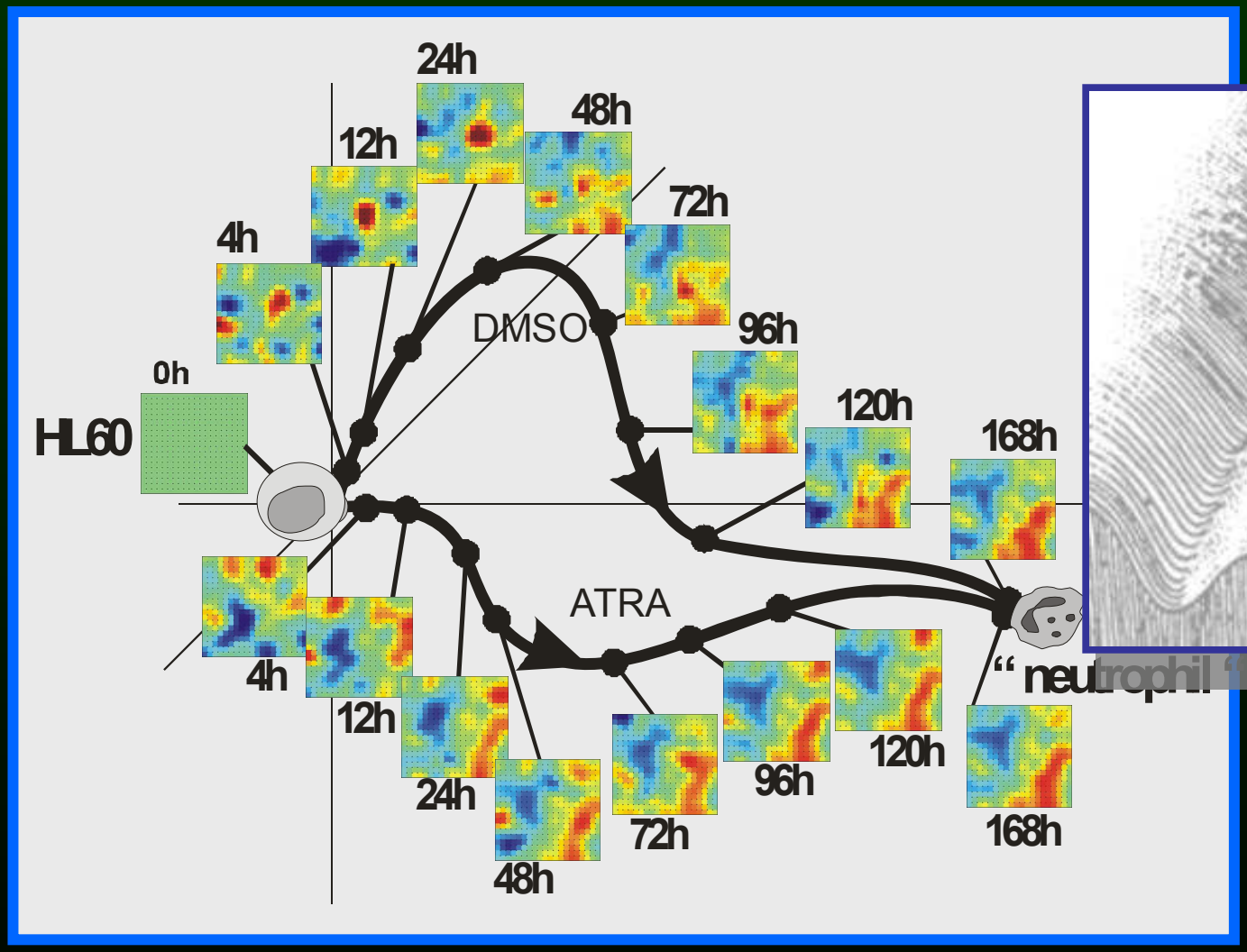
Two trajectories of HL-60 differentiation

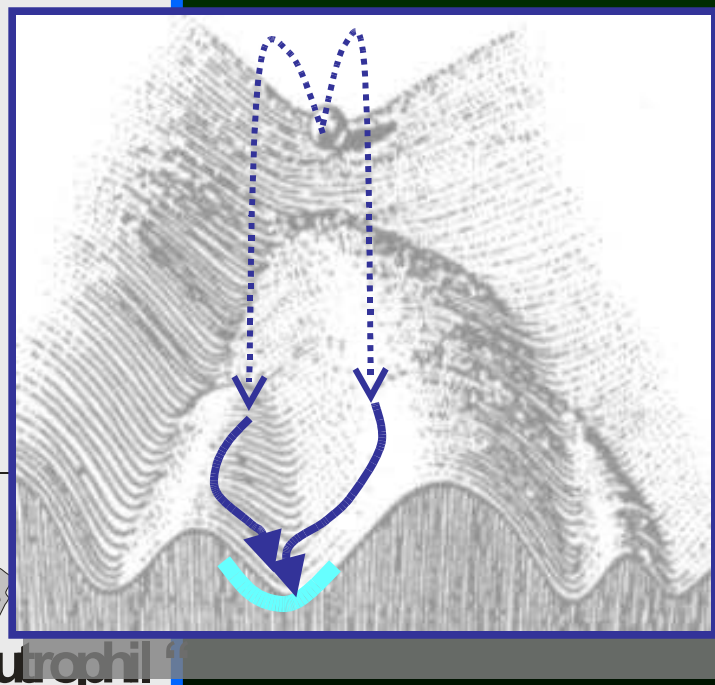
[For *N* = 2773 genes (= 72%) shared by “DMSO- and ATRA-neutrophils”]



Huang, Eichler, Bar-Yam, Ingber, PRL (2005)

Summary (1)





Convergence of two high-dimensional ($N \sim 2700$) trajectories as indication (necessary condition) for state space contraction

Biological observables to be explained

The four *D*'s:

DISCRETENESS ✓

DIFFERENTIATION ✓

DIVERSIFICATION

DIRECTIONALITY

Biological observables to be explained

The four *D*'s:

DISCRETENESS ✓

DIFFERENTIATION ✓

DIVERSIFICATION

DIRECTIONALITY

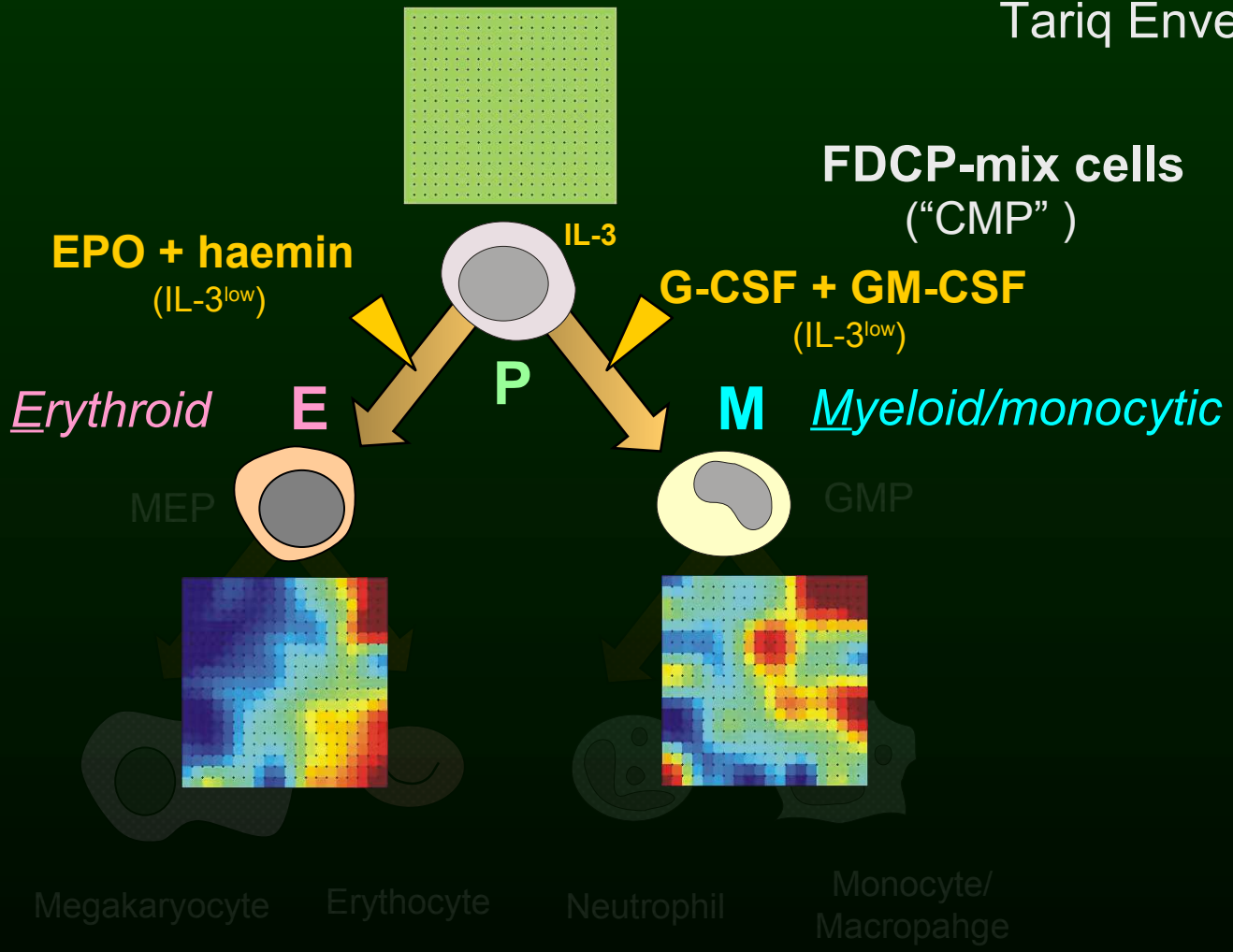
$$S^*_P(t_1) = [\, x_1(t_1),\, x_2(t_1) \ldots\, x_N(t_1)\,]$$

$$S^*_E(t_2) = [\, x_1(t_2),\, x_2(t_2),\, \ldots\, x_N(t_2)\,]$$

$$S_M(t_2) = [\, x_1(t_2),\, x_2(t_2),\, \ldots\, x_N(t_2)\,]$$

What is the very essence of “path separation” ?

Tariq Enver, Oxford

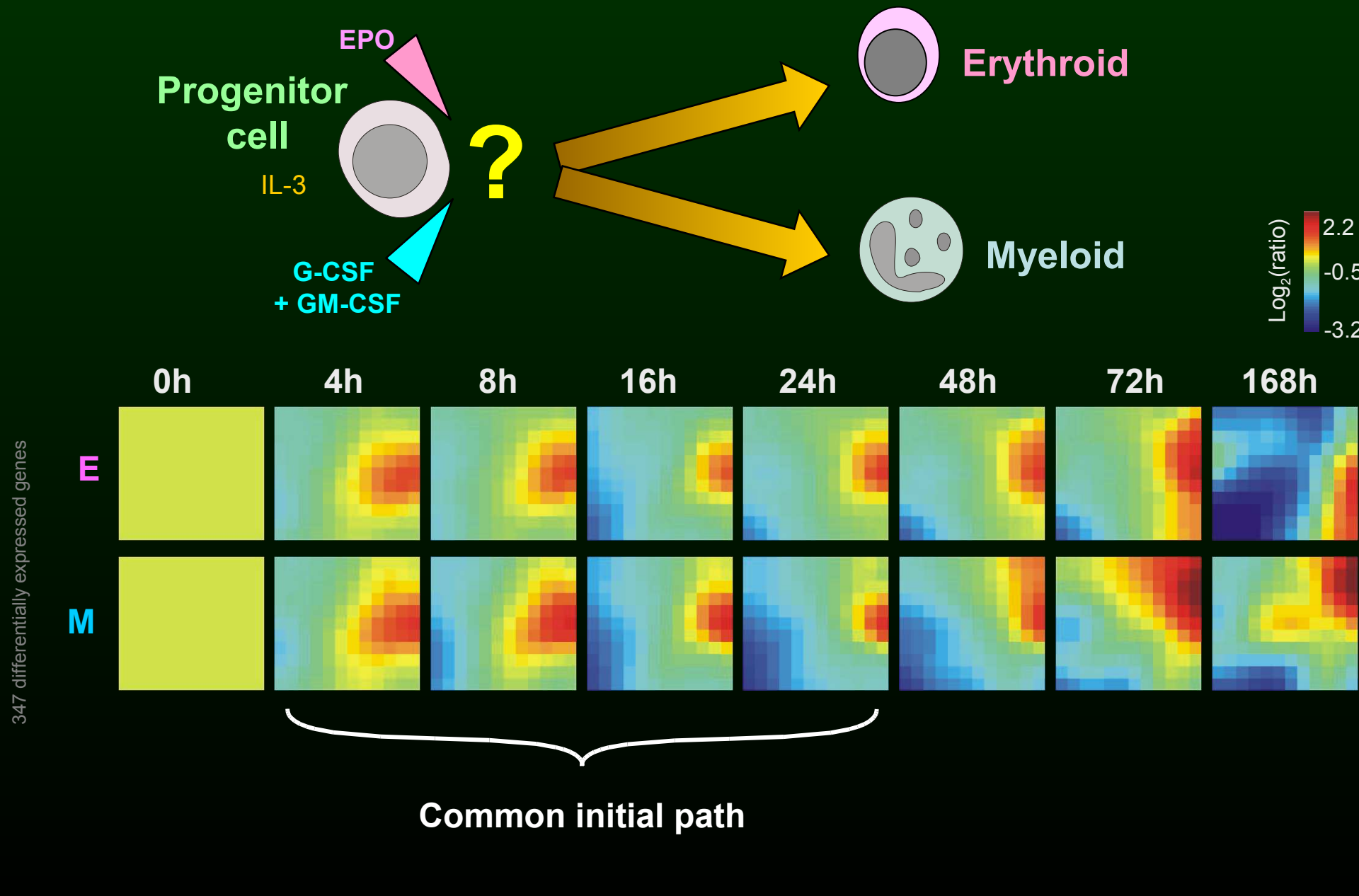


$S^*_P(t_1) = [\, x_1(t_1),\, x_2(t_1) \ldots\, x_N(t_1)\,]$

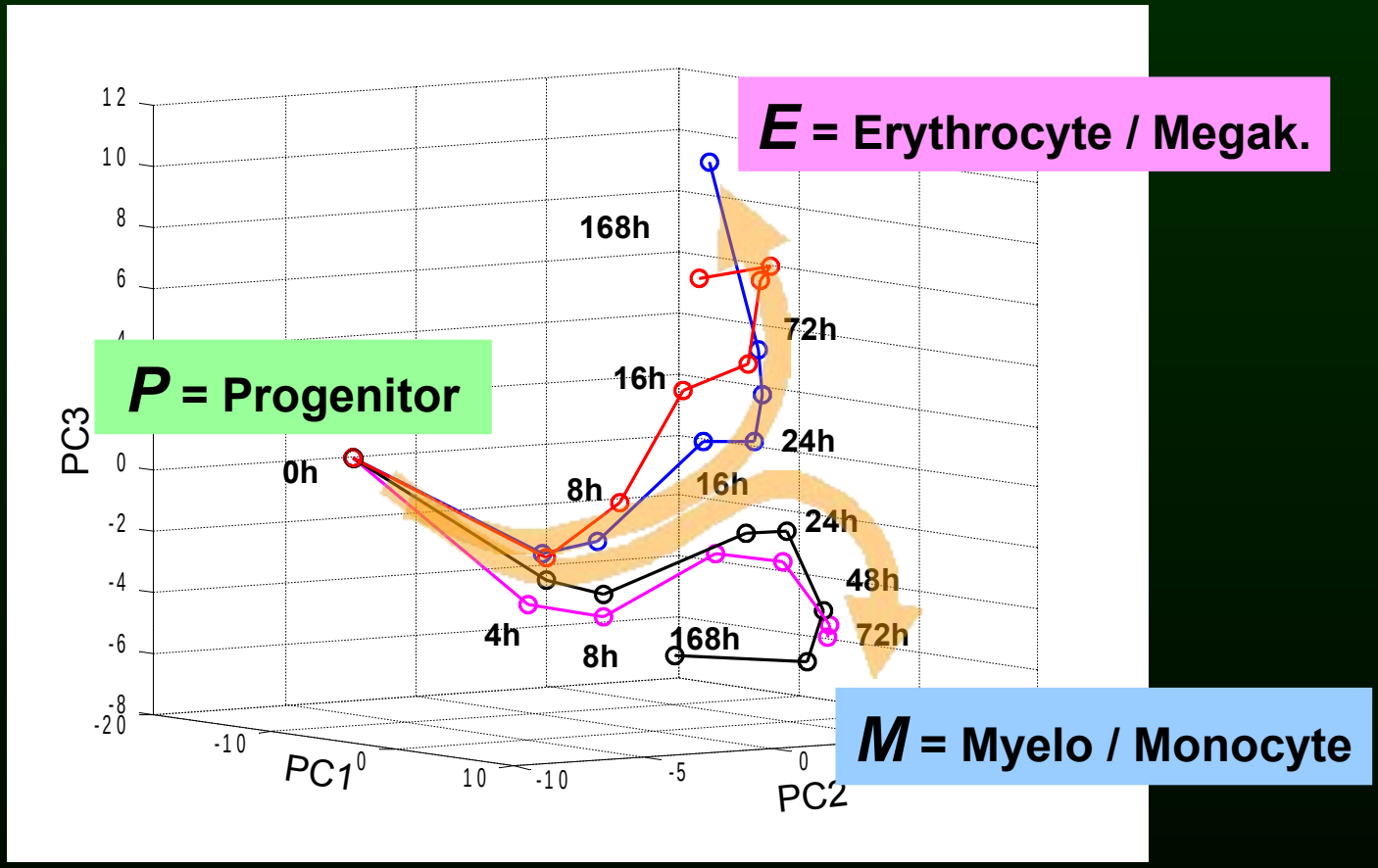
$S^*_E(t_2) = [\, x_1(t_2),\, x_2(t_2),\, \ldots\, x_N(t_2)\,]$

$S_M(t_2) = [\, x_1(t_2),\, x_2(t_2),\, \ldots\, x_N(t_2)\,]$

Gene expression profiling of cell fate decision

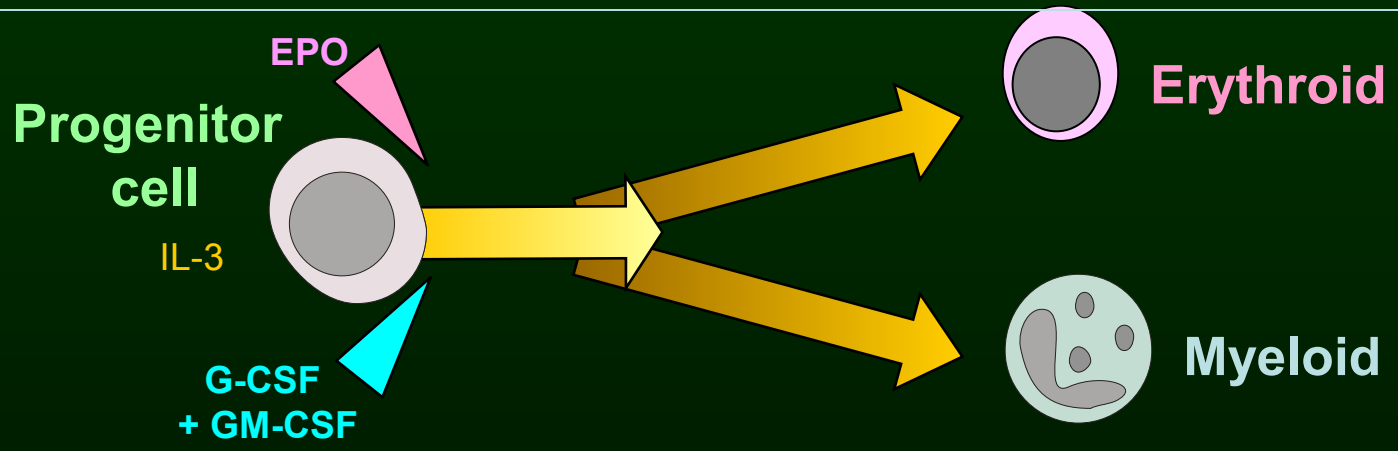


Principal component analysis



Hypothesis:

Fate commitment as two-phase process

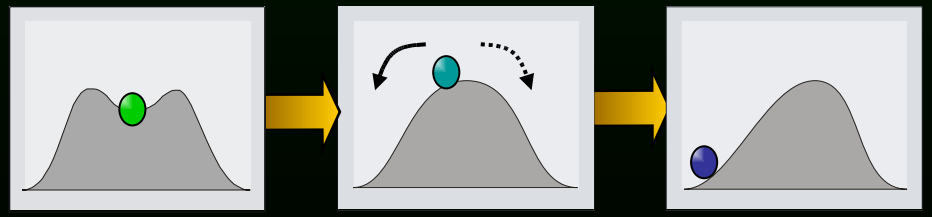


1. ***Destabilization*** of the multipotent progenitor cell ***P***

2. ***Decision***

- ***stochastic*** influences (noise)
- bias/small perturbation by "***deterministic***" signals

Attractor landscape view:
= "*bifurcation*" event

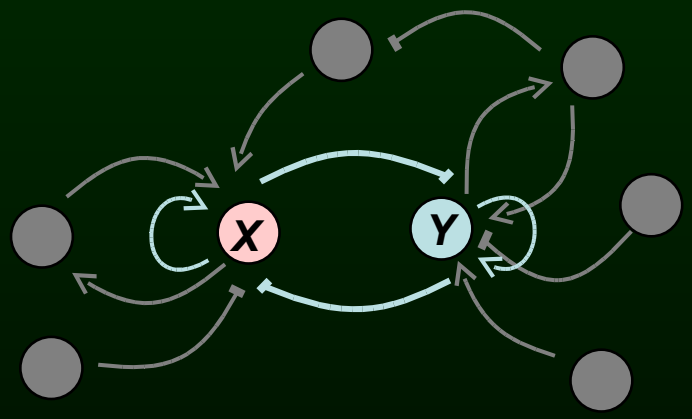


Low vs. High-dimensional system

Dilemma:

Tractable: ⟷ Reality:

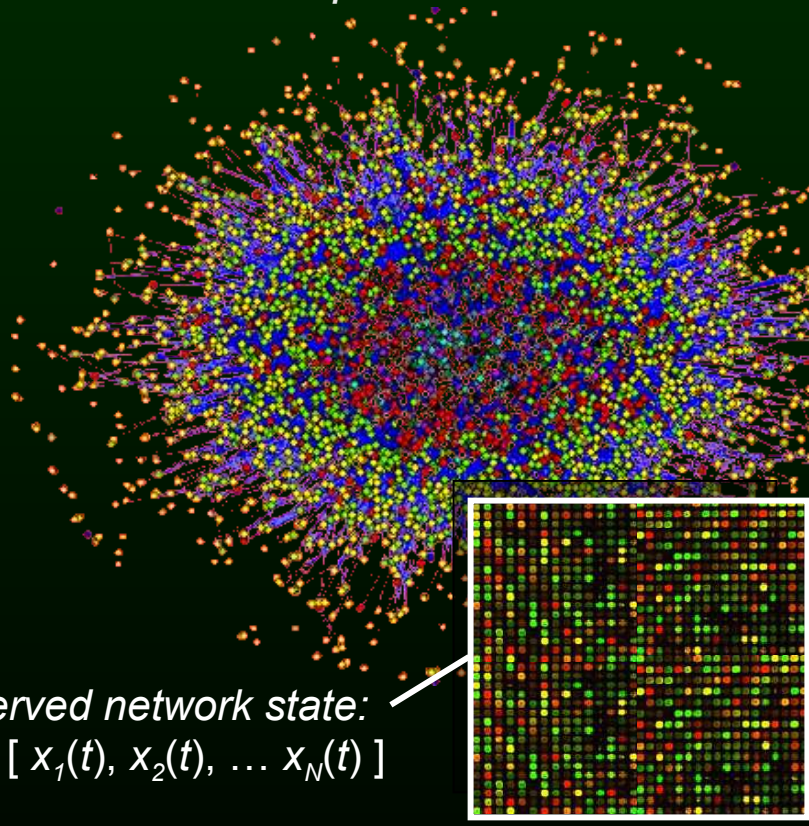
Small circuits



$$\frac{dX}{dt} = k_1 \frac{S_X{}^n}{S_X{}^n + Y^n} + k_2 \frac{X^n}{S_X{}^n + X^n} - k_3 X$$

$$\frac{dY}{dt} = k_3 \frac{S_Y{}^n}{S_Y{}^n + X^n} + k_5 \frac{Y^n}{S_Y{}^n + Y^n} - k_6 Y$$

Complex network



Observed network state:

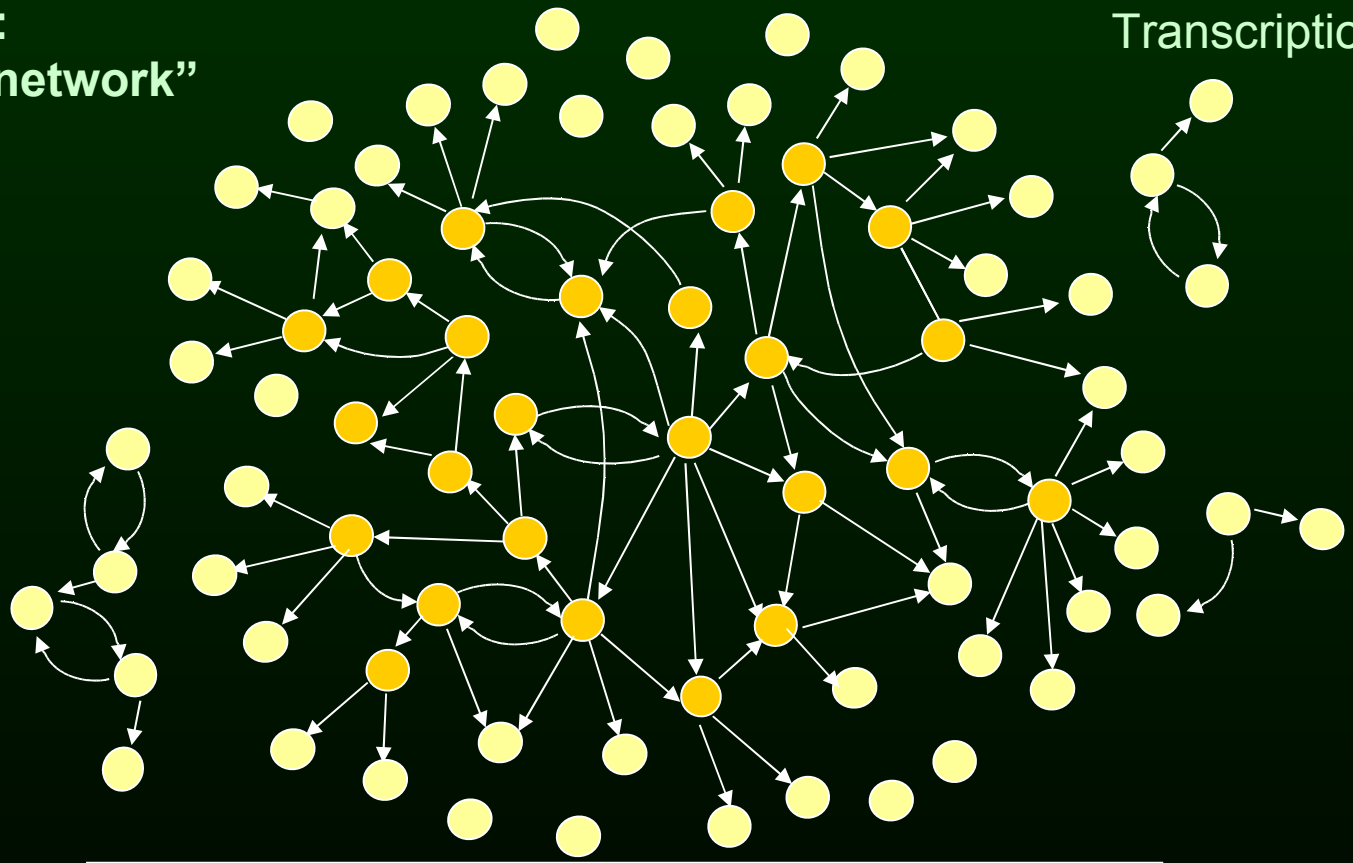
$\boldsymbol{S}(t) = [\, x_1(t), x_2(t), \ldots x_N(t) \,]$

A core network that determines dynamics

may be reasonably small

Kauffman:
"medusa network"

10-20 % of genome:
Transcription/signaling proteins



$$\boldsymbol{S}\,(t) = [\, x_1(t),\, x_2(t),\, ..\, x_i(t),\, x_k(t),\, x_{k+1}(t),.. \, x_N(t)\,]$$

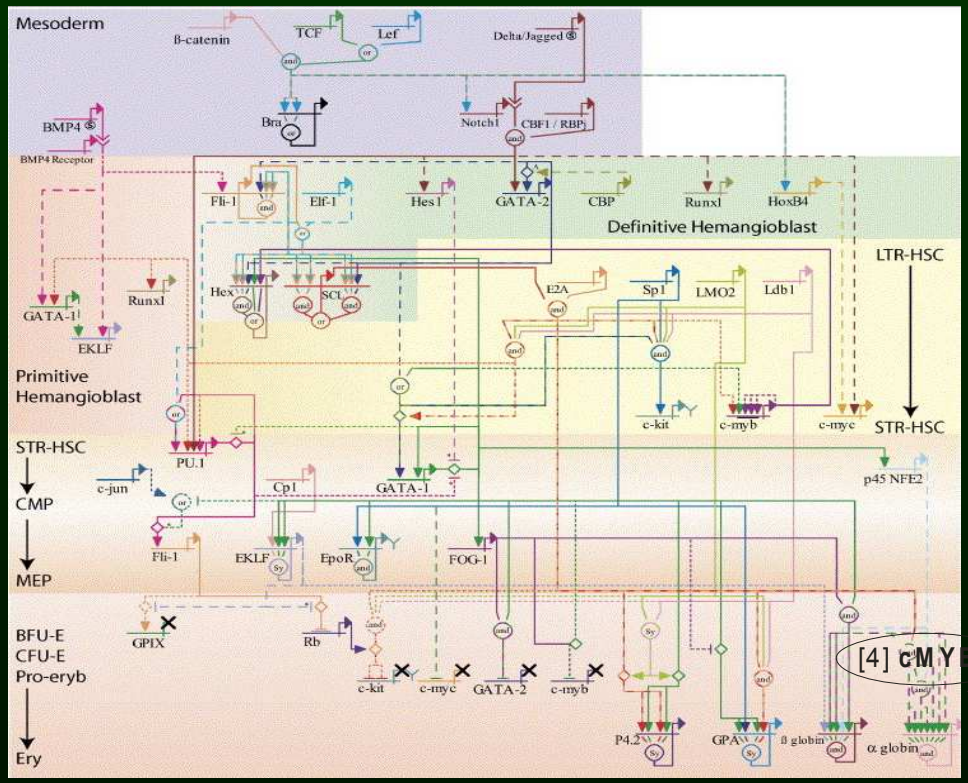
Determines Global dynamics of network

Core regulatory set
(Transcription factors, …)

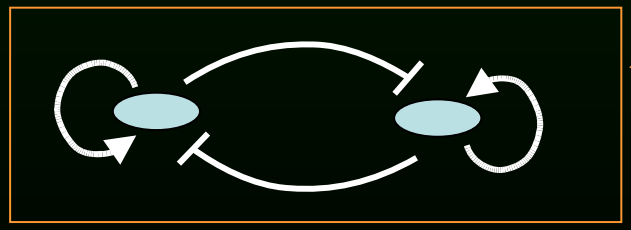
Peripheral genes
(non-regulatory effectors)

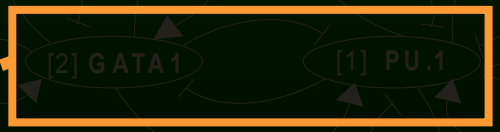
Additional degrees of freedom for environm. signals
(phenotype fine tuning)

comprehensive view:

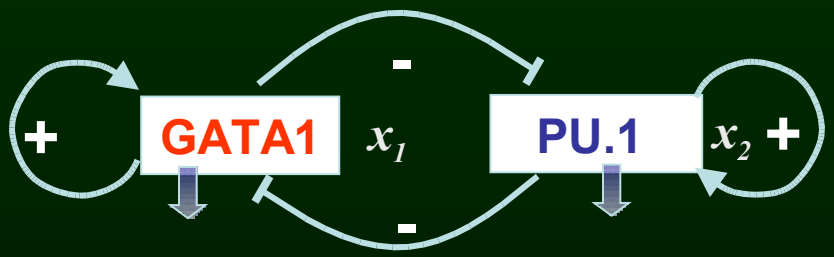


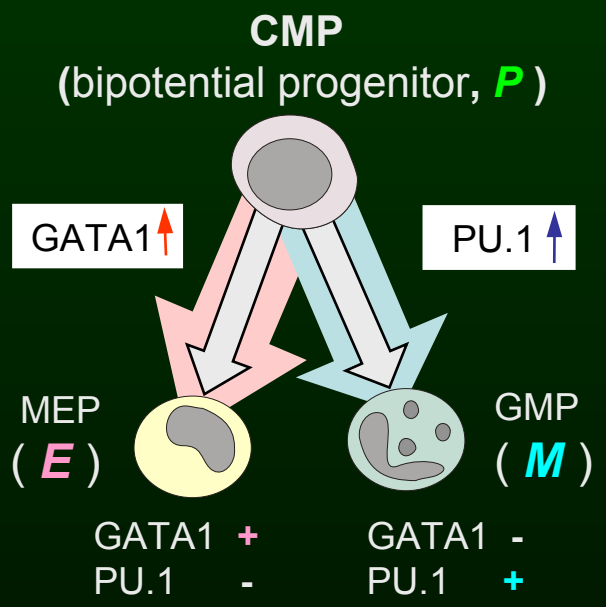
Transcription factors only ("medusa head")

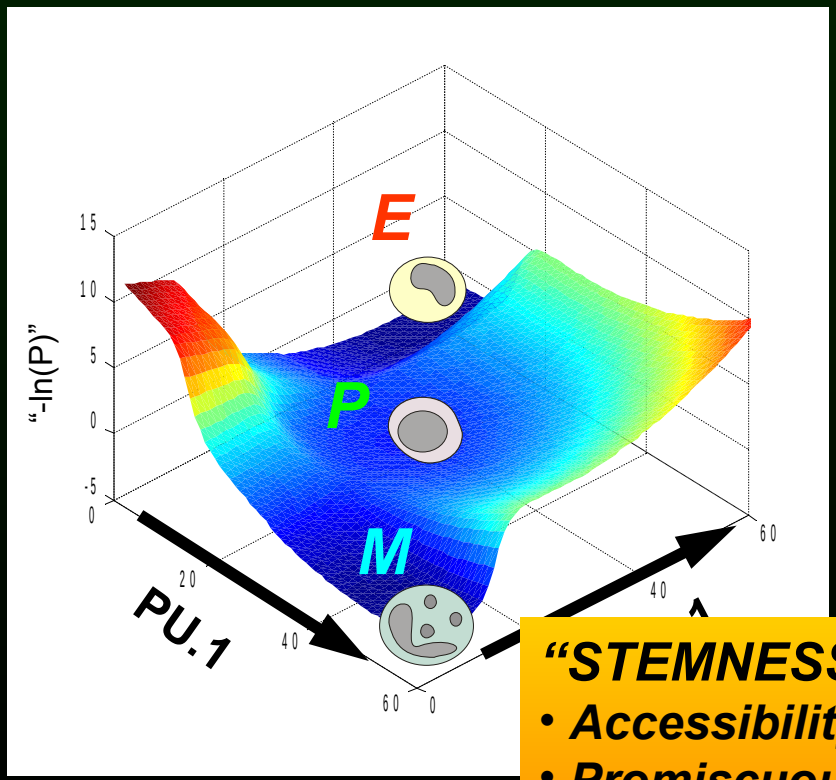


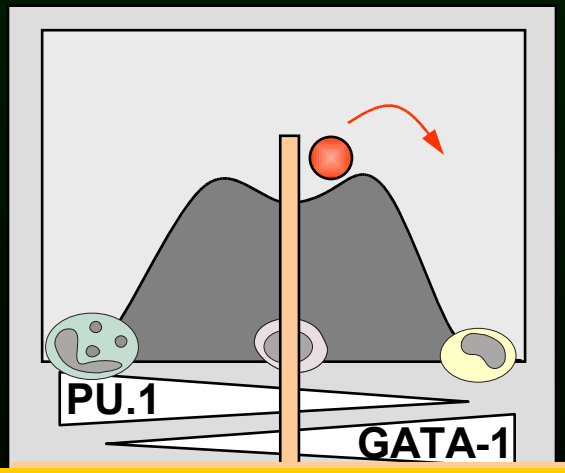


GATA1-PU.1 system in cell fate decision





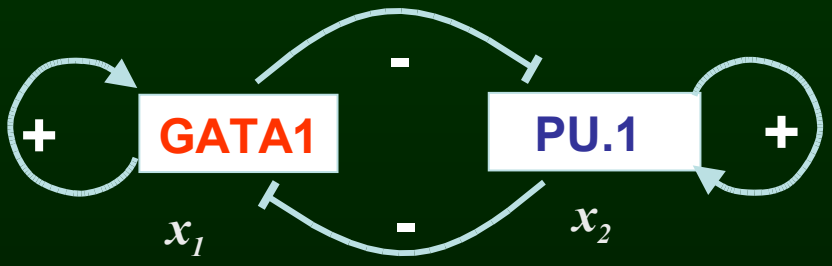




"STEMNESS" / MULTIPOTENCY

- ***Accessibility to multiple attractors !***
- ***Promiscuous ("preview") gene expression***

Pitchfork Bifurcation for $P \rightarrow (M,E)$

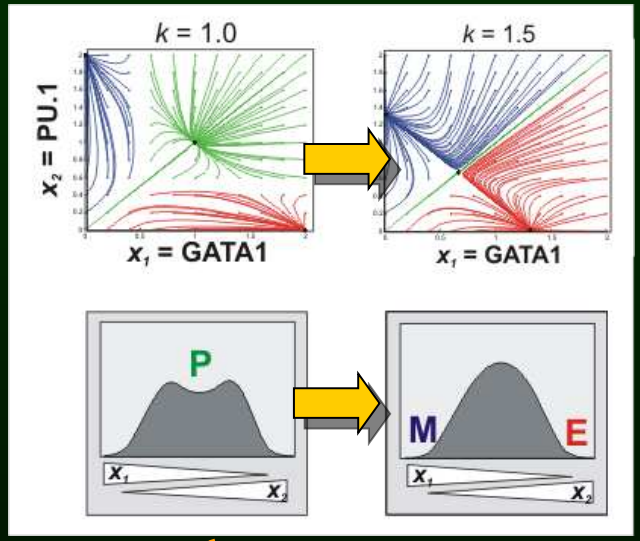


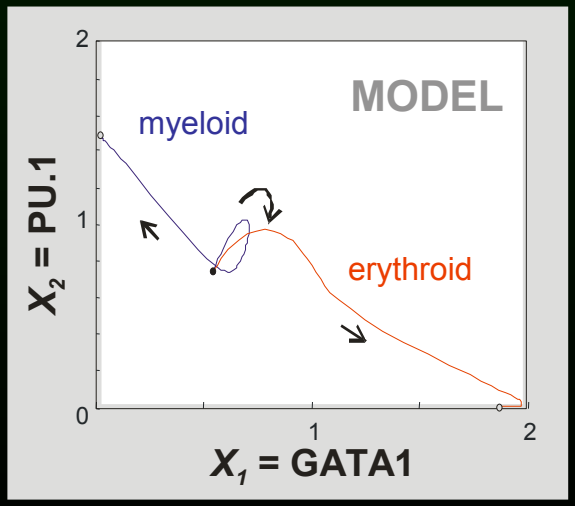
$$\frac{dx_1}{dt} = \frac{a_1{}^n x_1{}^n}{S^n + x_1{}^n} + \frac{b_1 S^n}{S^n + x_2{}^n} - k_1 x_1$$

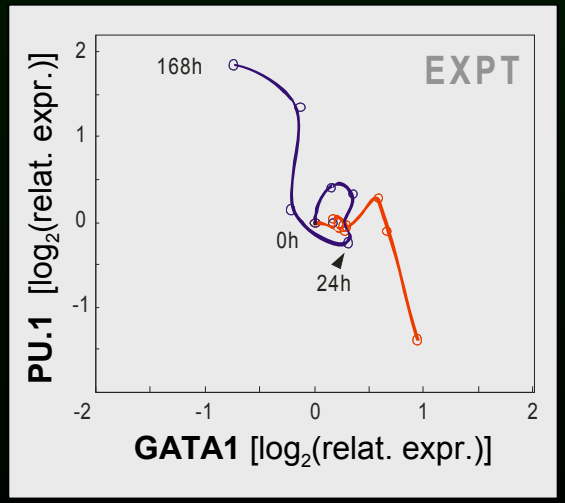
$$\frac{dx_2}{dt} = \frac{a_2{}^n x_2{}^n}{S^n + x_2{}^n} + \frac{b_2 S^n}{S^n + x_1{}^n} - k_2 x_2$$

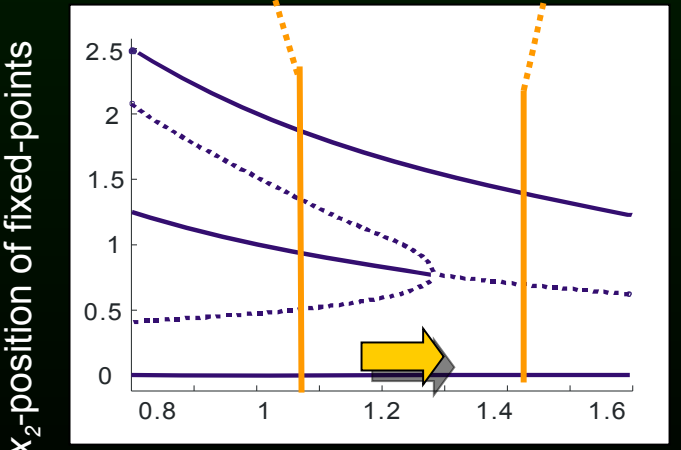
$$a_1 = a_1^0 \exp(-\lambda_1 t)$$

$$a_2 = a_2^0 \exp(-\lambda_2 t)$$

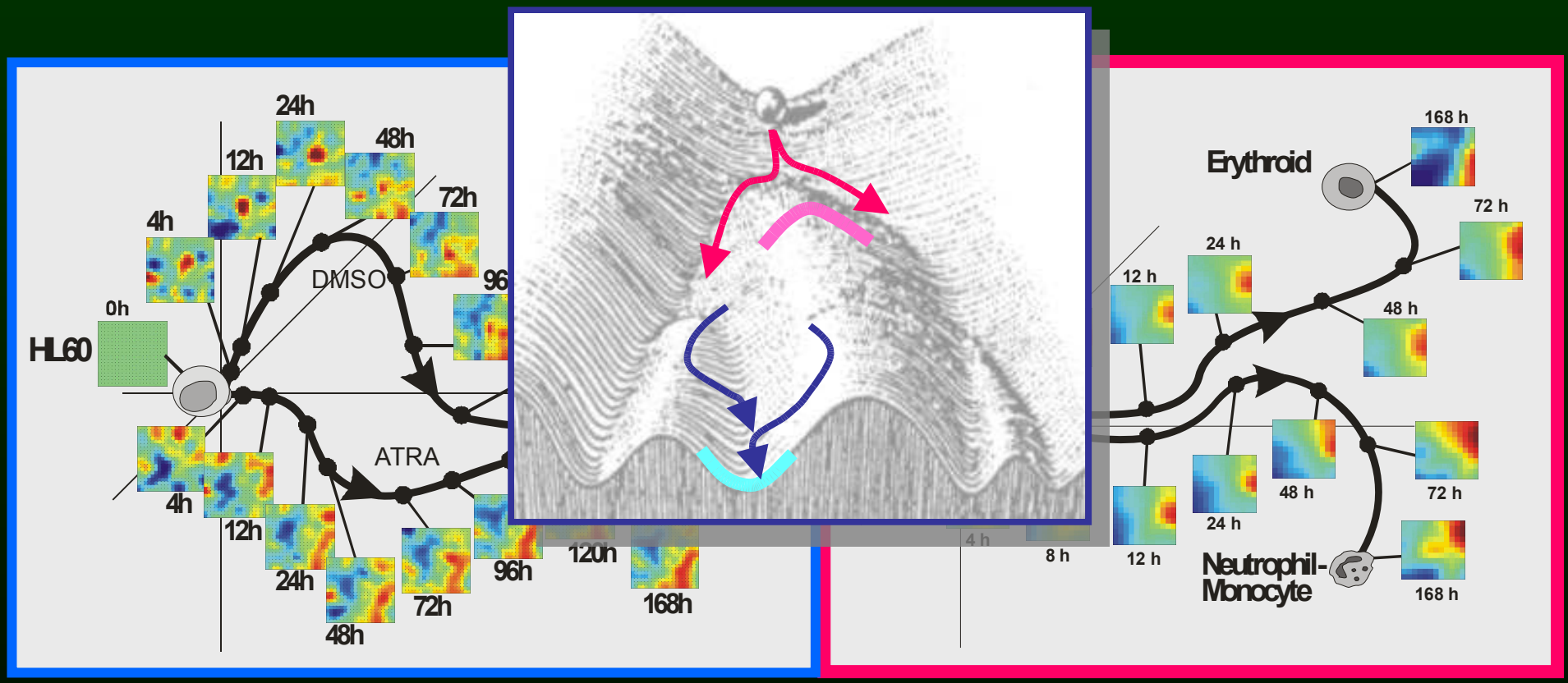








Summary (2)



Convergence

Divergence

→ **Cell fate as a high-dimensional attractor state (“ valley “)**

→ **Cell fates are separated by a repellor state (“ hill top“)**

Biological observables to be explained

The four *D*'s:

DISCRETENESS ✓

DIFFERENTIATION ✓

DIVERSIFICATION ✓

DIRECTIONALITY

$$S^*_P(t_1) = [\, x_1(t_1),\, x_2(t_1) \ldots\, x_N(t_1)\,]$$

$$S^*_E(t_2) = [\, x_1(t_2),\, x_2(t_2),\, \ldots\, x_N(t_2)\,]$$

$$S_M(t_2) = [\, x_1(t_2),\, x_2(t_2),\, \ldots\, x_N(t_2)\,]$$

Biological observables to be explained

The four *D*'s:

DISCRETENESS ✓

DIFFERENTIATION ✓

DIVERSIFICATION ✓

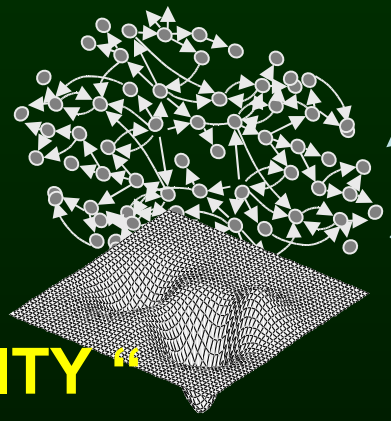
DIRECTIONALITY ← *Current work…*

So far: every regulatory event is reversible.

Hypothesis:

Directionality emerges at network level + probabilistic events

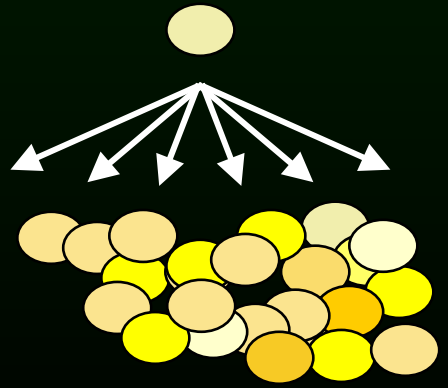
Single cell level analysis of multicellularity



One genome

?

Nominally uniform clone



Inter-type Diversity

(Stable cell types, subtypes)

“ MACRO-HETEROGENEITY “

“ MICRO-HETEROGENEITY “

Intra-type Heterogeneity

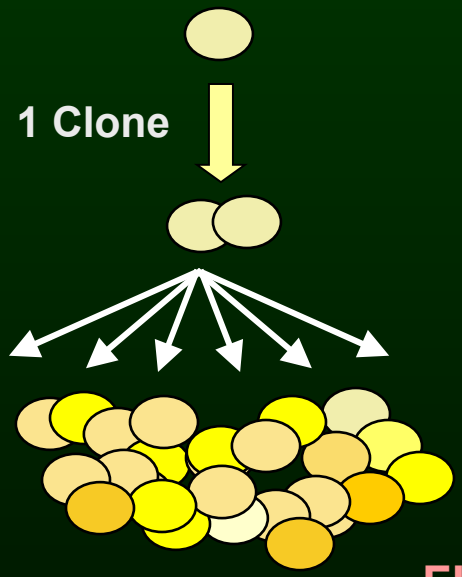
(Stochastic epigenetic population heterogeneity)

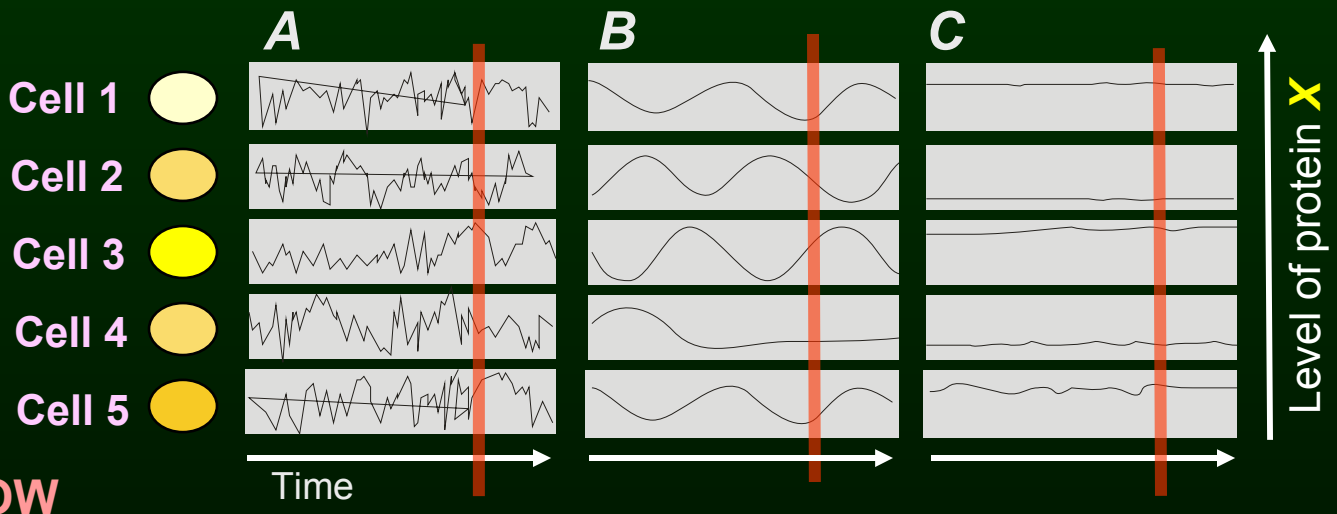
Old idea: Spudich & Koshland (1976); …

Recent: Singer (2001); Leibler (2004); Cluzel (2005); etc

→ “ Gene expression noise ”

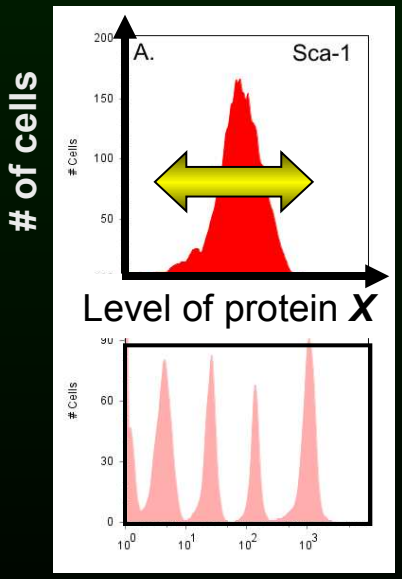
Basis for intra-type heterogeneity





FLOW CYTOMETRY

Snapshot t_1 Snapshot t_1 Snapshot t_1

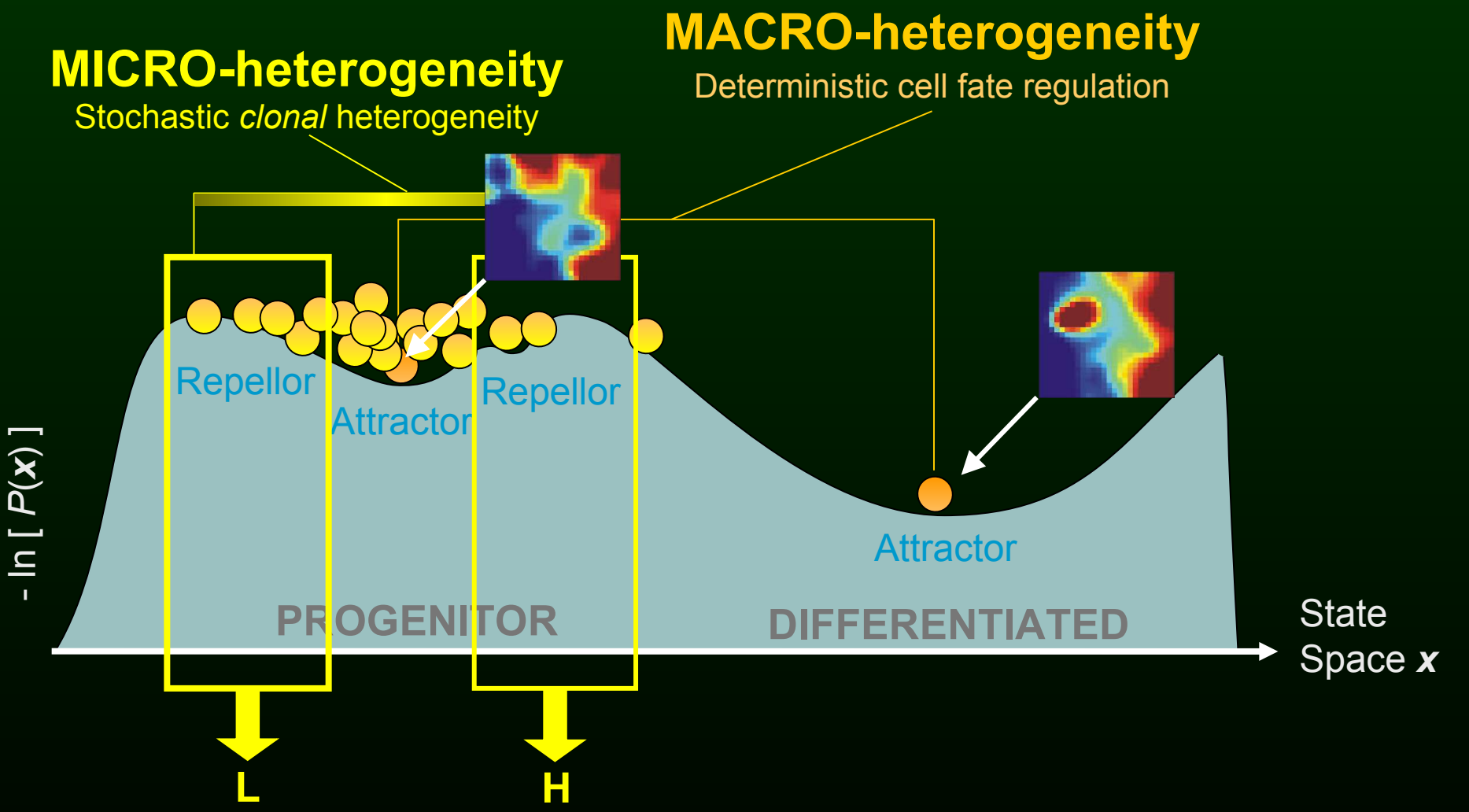


10-1000 fold range
Of protein ***X***

If dynamics is slow:
Is the typical time scale of cell regulatory processes < the 'ergodic regime' ??

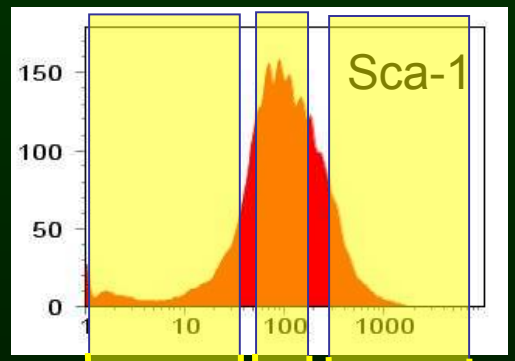
Then, not just *ensemble average* but *outliers* can also determine macroscopic biology (different from thermodynamics!)

Exploring MICRO-HETEROGENEITY

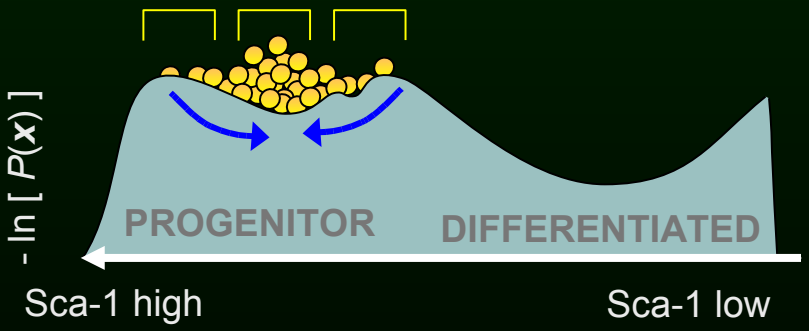


- **Kinetics ?** **Slow, with distinct cell behaviors - Or just noisy fluctuation ?**

Slow relaxation kinetics of Sca-1 outliers in clonal population

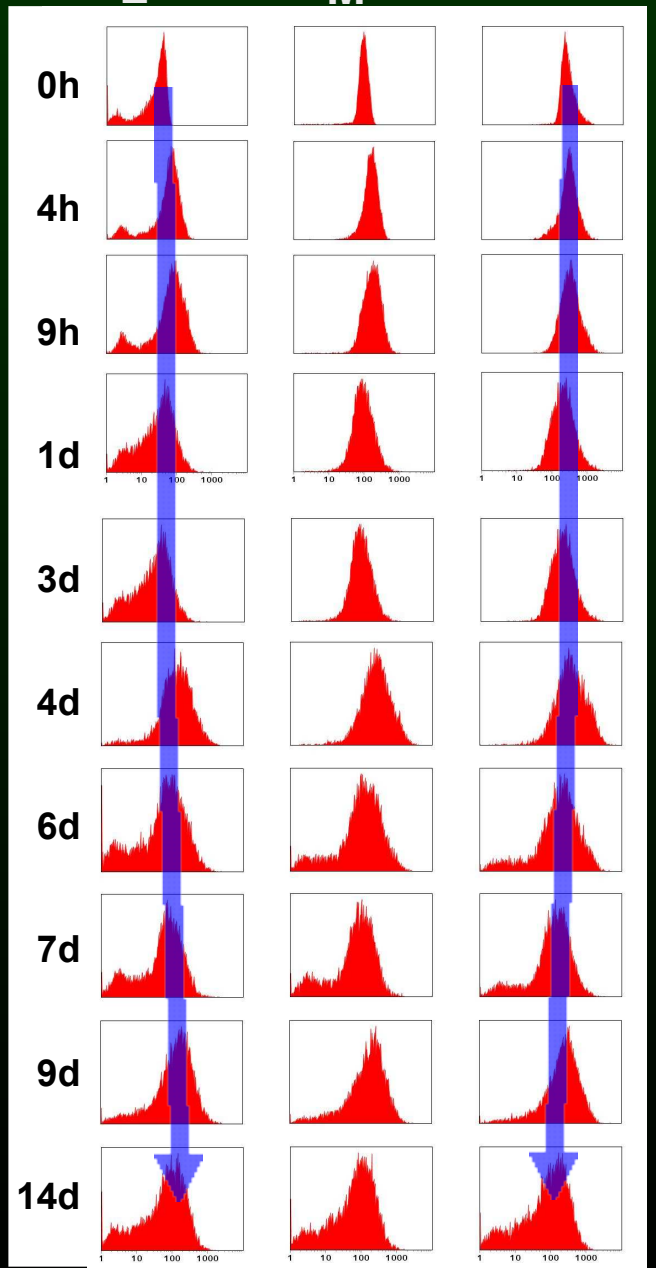


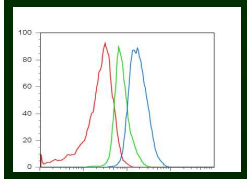
Clonal population
Of EML cells

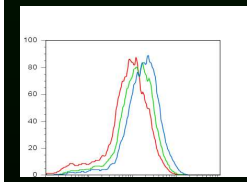


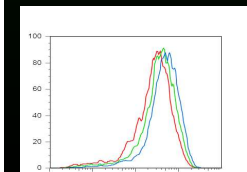
Sorted fraction: L M H

Time (~ 10 cell generations)



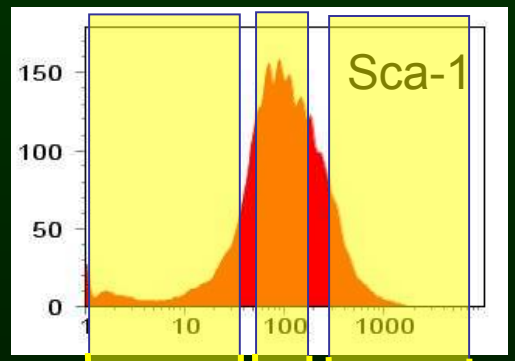




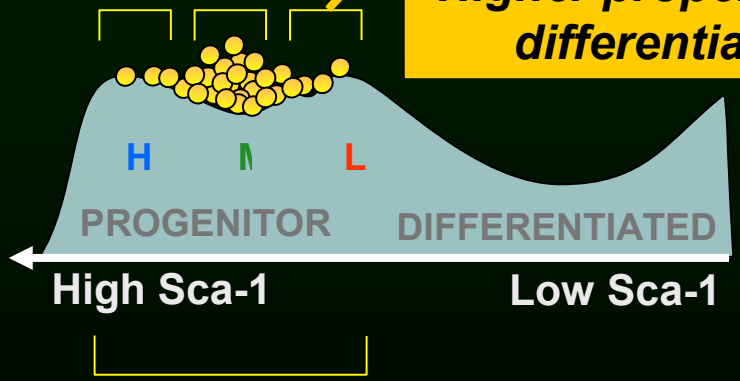


- **Population dispersion is caused by slow relaxing individuality**
- **Parental distribution is established → attractor state**

Micro-heterogeneity of progenitors and differentiation rate



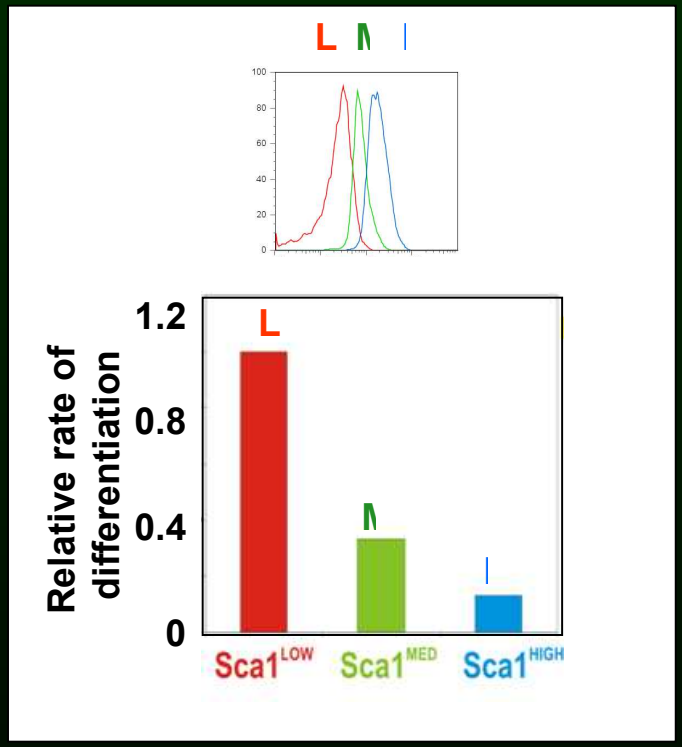
L M H



Higher propensity to differentiate?

= 1 nominally uniform population in common biology

+ Erythropoietin (EPO)
→ measure rate of differentiation



Cells that "happen" to express low Sca-1, are more prone to differentiate.
Stochastic, epigenetic cell heterogeneity has biological implications.

More details
to follow tomorrow

Biological observables to be explained

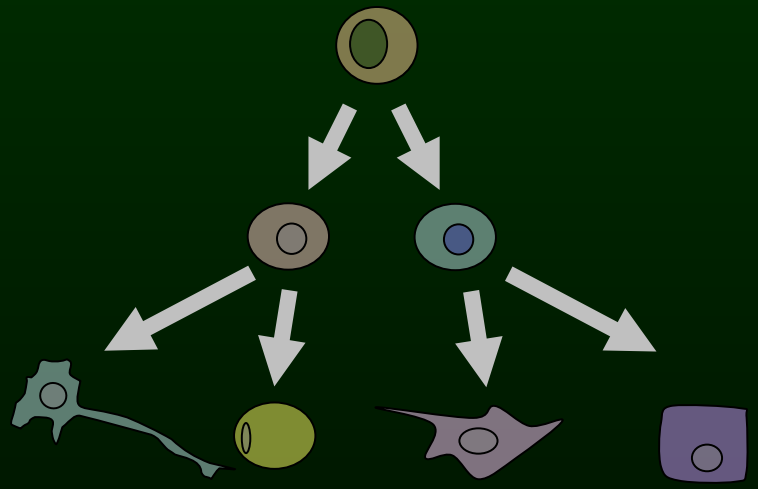
The four *D*'s:

DISCRETENESS

DIFFERENTIATION

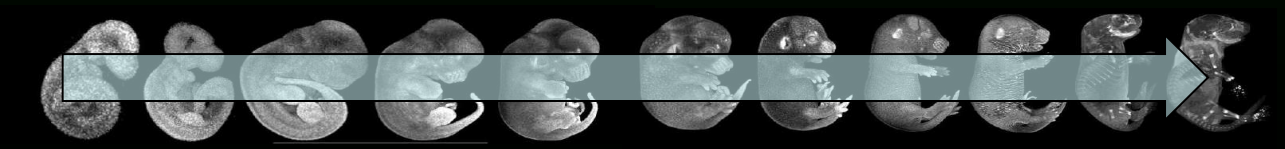
DIVERSIFICATION

DIRECTIONALITY



$S_1 = [x_1, x_2, .., x_N]$

$S_1 = [x_1, x_2, .., x_N]$ $S_2 = [x_1, x_2, .., x_N]$



For biologists: a new “biological observable”

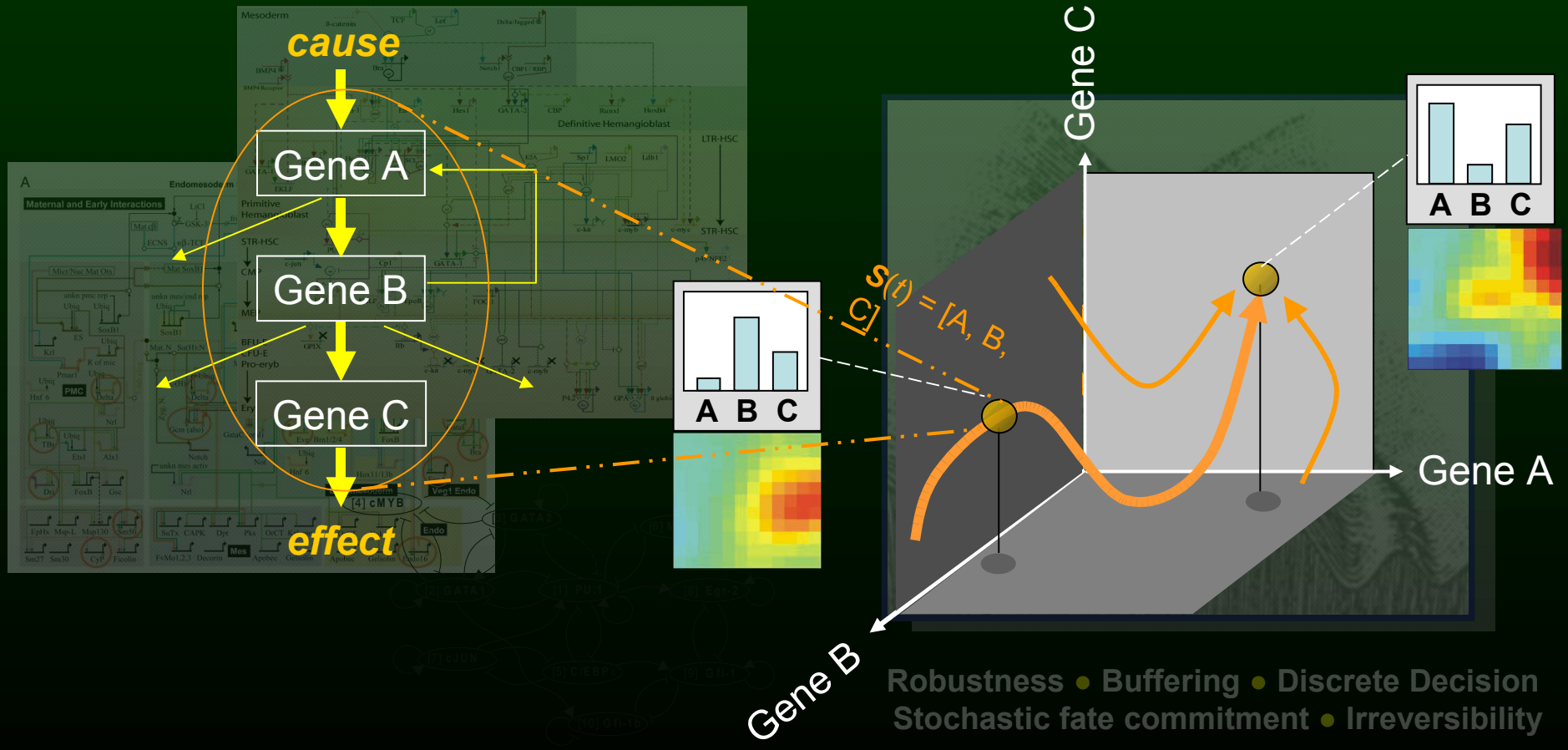
Pathway

in network diagrams

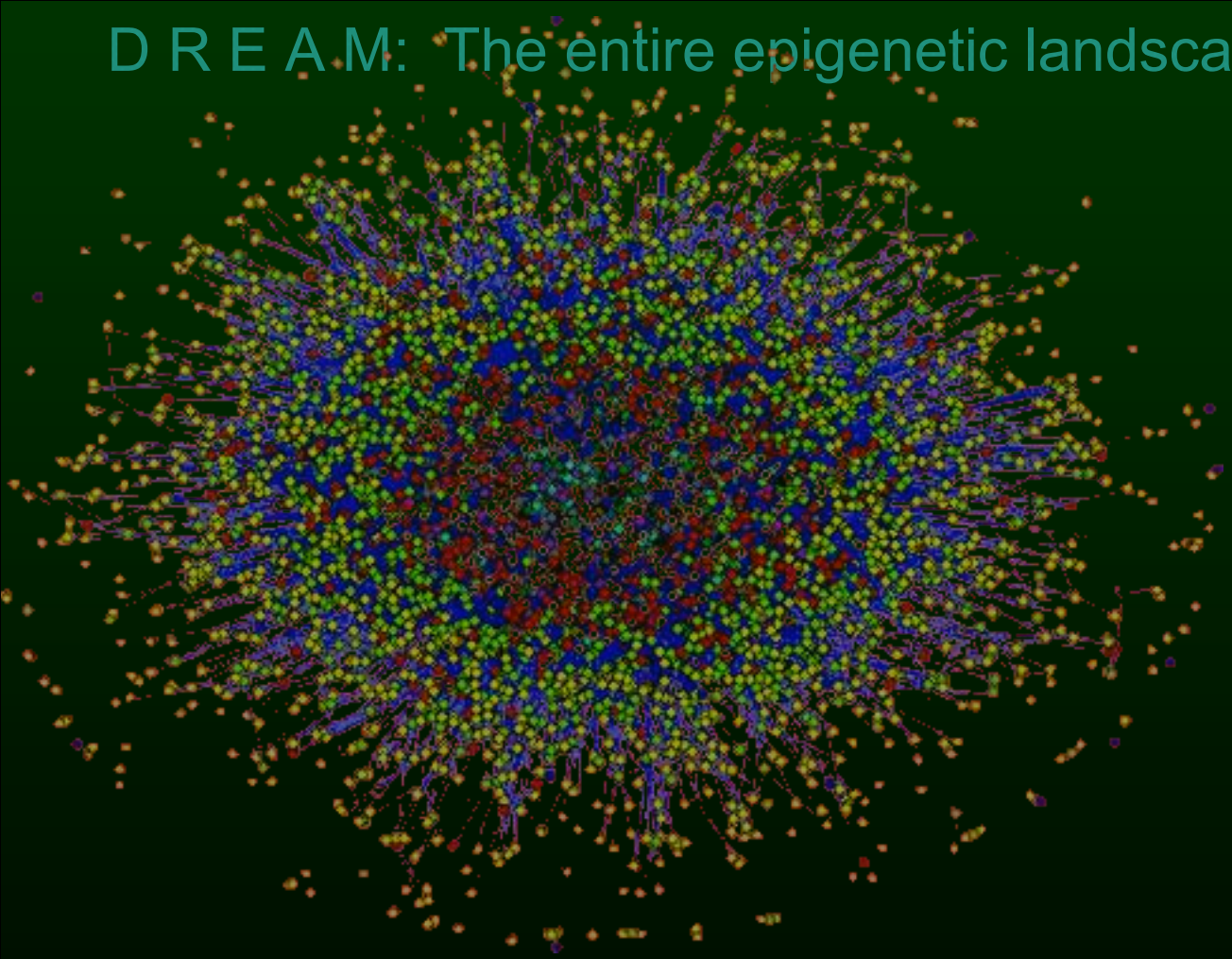
Path

in state space

= developmental path = trajectory

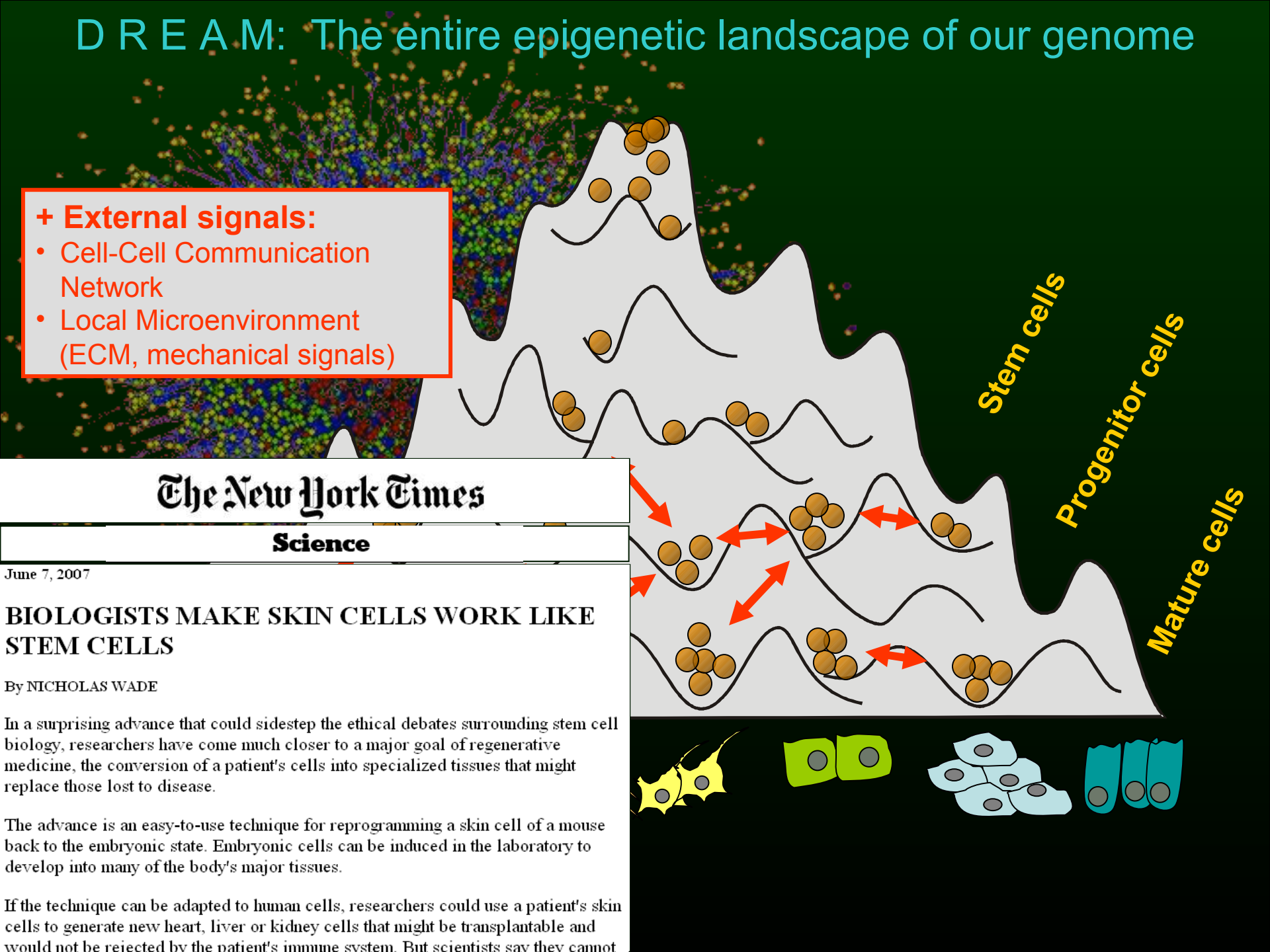


D R E A M: The entire epigenetic landscape of our genome



D R E A M: The entire epigenetic landscape of our genome

+ External signals:
- Cell-Cell Communication Network
- Local Microenvironment (ECM, mechanical signals)



The New York Times

Science

June 7, 2007

BIOLOGISTS MAKE SKIN CELLS WORK LIKE STEM CELLS

By NICHOLAS WADE

In a surprising advance that could sidestep the ethical debates surrounding stem cell biology, researchers have come much closer to a major goal of regenerative medicine, the conversion of a patient's cells into specialized tissues that might replace those lost to disease.

The advance is an easy-to-use technique for reprogramming a skin cell of a mouse back to the embryonic state. Embryonic cells can be induced in the laboratory to develop into many of the body's major tissues.

If the technique can be adapted to human cells, researchers could use a patient's skin cells to generate new heart, liver or kidney cells that might be transplantable and would not be rejected by the patient's immune system. But scientists say they cannot

Acknowledgement

BIOLOGY

BIOINFORMATICS
MATH PHYSICS

Children's Hospital,
Harvard Medical School, Boston

Hannah Chang
Philmo Oh
Gabe Eichler
Yuchun Guo

Don Ingber

Univ. Oxford

Tariq Enver
Shamit Soneji

New England Inst of Complexity

Yaneer Bar-Yam

Imperial College, London

Martin Hemberg