Bioinformatics 3

V10 –
Simulating the Dynamics of Gene Regulatory Networks by Boolean Networks

Fri, Nov 27, 2015

Quorum sensing of *Vibrio fischeri*

This luminescent bacterium exists in small amounts in the ocean and in large amount in isolated areas such as the light organs of squid.

When in small concentrations of cells, *V. fischeri* does not give off light,
but in high cell density these bacteria emit a blue-green light.

This cell density-dependent control of gene expression is activated by auto-induction that involves the coupling of a transcriptional activator protein with a signal molecule (autoinducer) that is released by the bacteria into its surrounding environment.

In the ocean, the population density of *V. fischeri* is only about 10^2 cells/ml.

Exporting the autoinducer from the bacteria into this low concentration
of cells is not enough to cause the luminescence genes to be activated.

However, inside the light organ of a squid for example,
the cell concentration is about 10^{10} cells/ml.

At such high concentrations, the autoinducer causes the bacteria to emit light

https://www.bio.cmu.edu/courses/03441/TermPapers/99TermPapers/Quorum/vibrio_fischeri.html

Quorum sensing of *Vibrio fischeri*

V. fischeri has a microbial **symbiotic relationship** with the squid *Euprymna scolopes*. The light organ of the squid provides the bacteria all of the nutrients that they need to survive. The squid benefits from the bacteria's quorum sensing and bioluminescence abilities.

During the day, the squid keeps the bacteria at lower concentrations by expelling some of them into the ocean during regular intervals.
At night however, the bacteria are allowed to accumulate to about 10^{10} cells/ml so that they will emit blue-green light.

This is perfect for the squid because it is a night feeder.

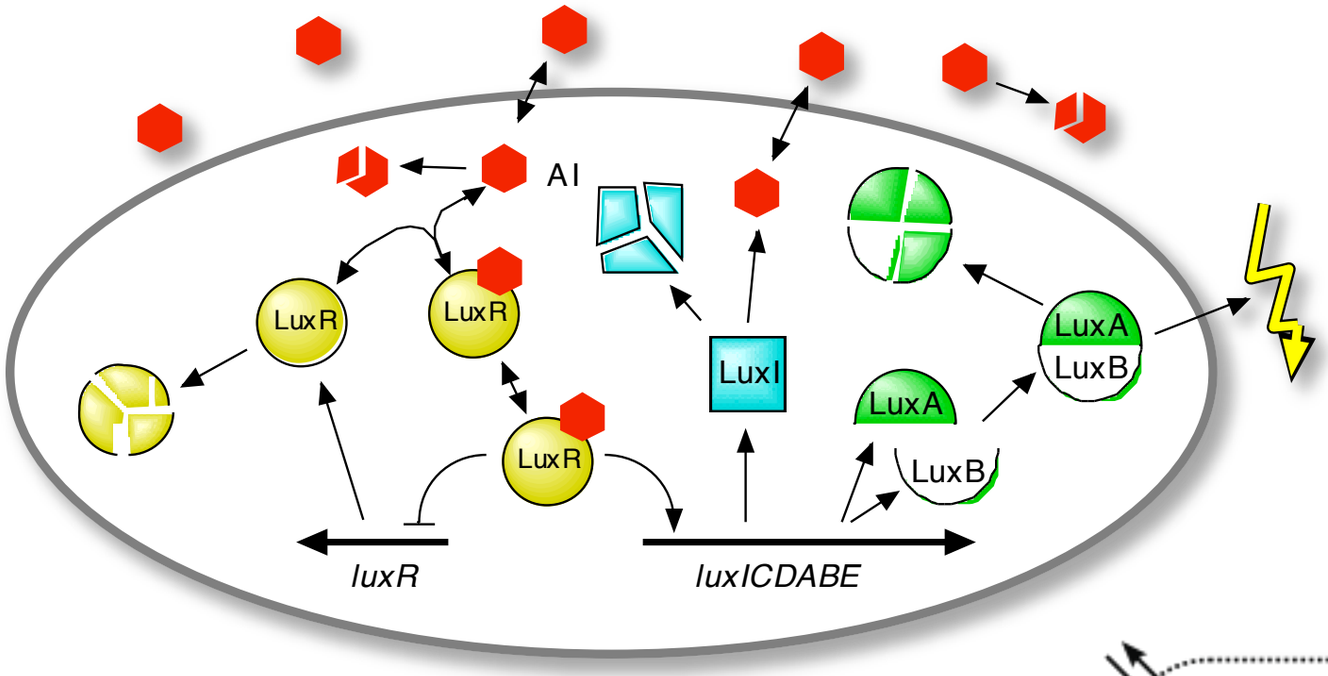
In the **moonlight**, the swimming squid would normally cast a **shadow** beneath itself making it a perfect target for squid-eating organisms.

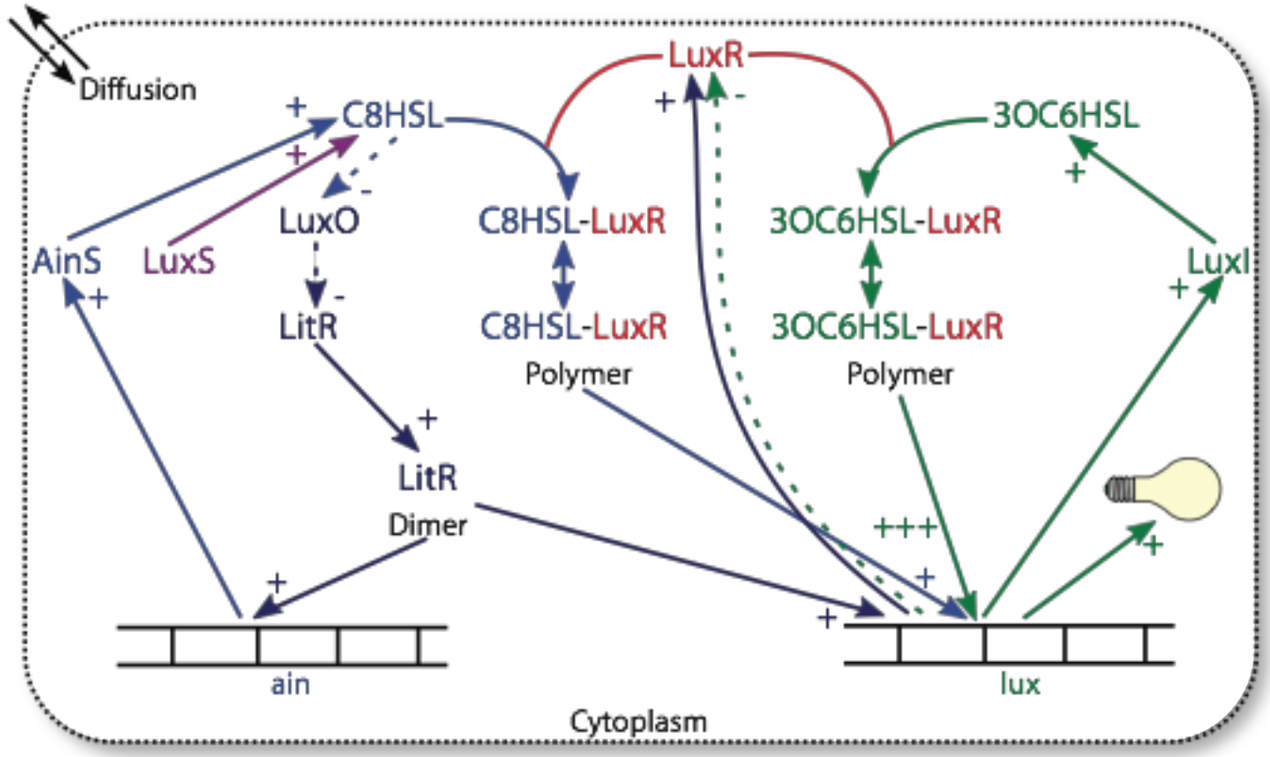
However, the bacterial glow will counter the shadowing effect the moon makes and mask the squid from its predators.

In the morning, the squid expels some bacteria into the ocean to a concentration where they will not generate light anymore so as to conserve energy.

https://www.bio.cmu.edu/courses/03441/TermPapers/99TermPapers/Quorum/vibrio_fischeri.html

Quorum sensing of *Vibrio fischeri*





Boolean Networks

"Blackboard explanations" often formulated as **conditional transitions**

- "If LuxI is present, then AI will be produced…"
- "If there is AI and there's no LuxR:AI bound to the genome, then LuxR will be expressed and complexes can form…"
- "If LuxR:AI is bound to the genome, then LuxI is expressed…"

Simplified mathematical **description** of the dependencies:

Densities of the species	<=>	discrete states: on/off, 1/0
Network of dependencies	<=>	condition tables
Progress in time	<=>	discrete propagation steps

Boolean Networks II

State of the system: described by **vector** of **discrete** values

$$S_i = \{0, 1, 1, 0, 0, 1, \ldots\}$$

$$S_i = \{x_1(i),\ x_2(i),\ x_3(i),\ \ldots\}$$

fixed number of species with **finite number** of states each

→ finite number of system states

→ periodic trajectories

→ **periodic** sequence of states = **attractor**

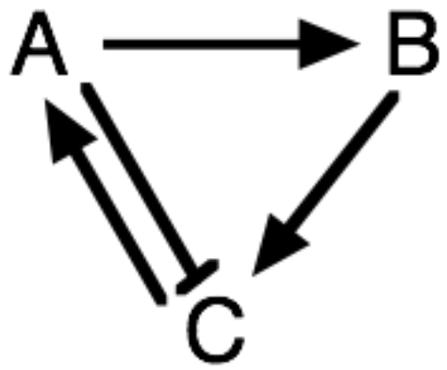
→ all states leading to an attractor = **basin of attraction**

Propagation:

$$S_{i+1} = \{x_1(i+1),\ x_2(i+1),\ x_3(i+1),\ \ldots\}$$

$$x_1(i+1) = f_1(x_1(i), x_2(i), x_3(i), \ldots)$$ with f_i given by condition tables

A Small Example



State vector $S = \{A, B, C\} \rightarrow$ 8 possible states

Conditional evolution:

A is on if C is on

A_{i+1}	C_i
0	0
1	1

A activates B

B_{i+1}	A_i
0	0
1	1

C is on if (B is on && A is off)

C_{i+1}	A_i	B_i
0	0	0
1	0	1
0	1	0
0	1	1

assume here that inhibition through A is stronger than activation via B

Start from {A, B, C} = {1, 0, 0}

#	S_i	A	B	C
0	S_0	1	0	0
1	S_1	0	1	0
2	S_2	0	0	1
3	$S_3 = S_0$	1	0	0

periodic orbit of length 3

Test the Other States

Test the other states

#	A	B	C
0	1	1	1
1	1	1	0
2	0	1	0
3	0	0	1
4	1	0	0
5	0	1	0

A_{i+1}	C_i
0	0
1	1

B_{i+1}	A_i
0	0
1	1

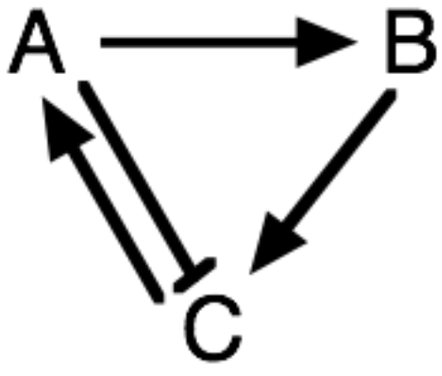
C_{i+1}	A_i	B_i
0	0	0
1	0	1
0	1	0
0	1	1

#	A	B	C
0	1	0	1
1	1	1	0

#	A	B	C
0	0	1	1
1	1	0	1

Same attractor as before:
100 → 010 → 001 → 100

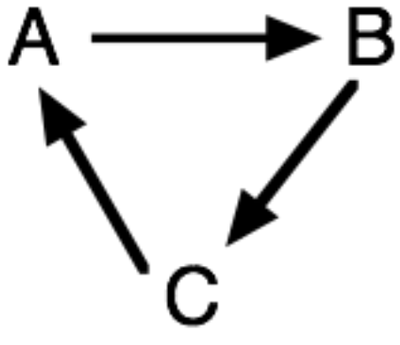
also reached from:
110, 111, 101, 011



#	A	B	C
0	0	0	0
1	0	0	0

→ **Either all off or stable oscillations**

A Knock-out Mutant



A_{i+1}	C_i
0	0
1	1

B_{i+1}	A_i
0	0
1	1

C_{i+1}	B_i
0	0
1	1

Attractors:

#	A	B	C
0	1	0	0
1	0	1	0
2	0	0	1
3	1	0	0

#	A	B	C
0	1	1	0
1	0	1	1
2	1	0	1
3	1	1	0

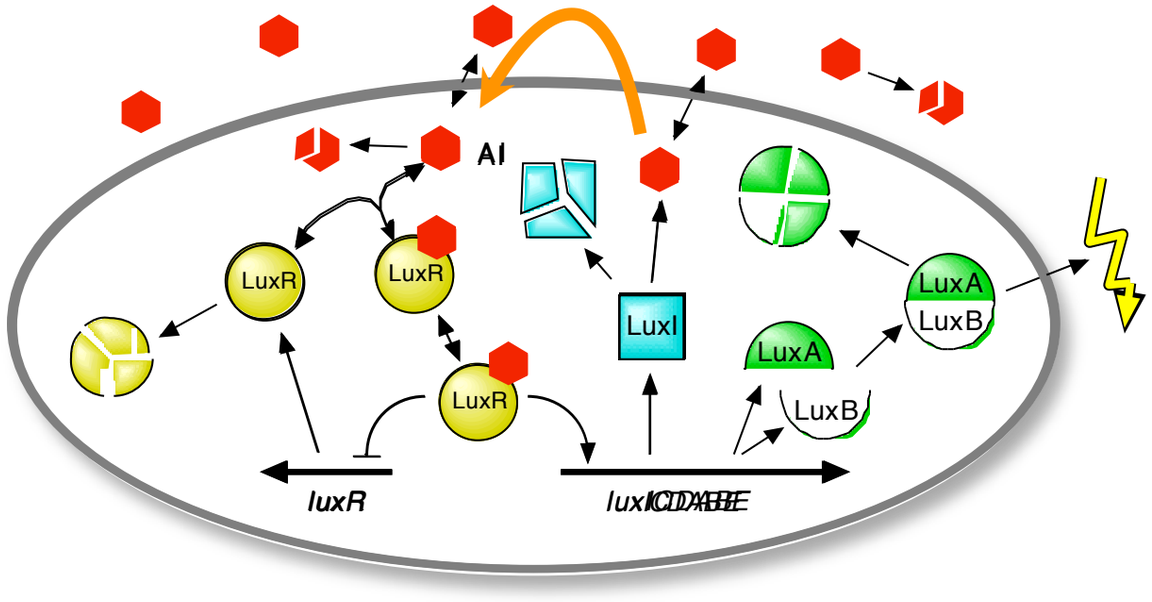
#	A	B	C
0	1	1	1
1	1	1	1

#	A	B	C
0	0	0	0
1	0	0	0

no feedback

→ no stabilization, network just "rotates"

Boolean Network of QS



Minimum set of species:

LuxR, AI, LuxR:AI, LuxR:AI:genome, LuxI

Here: Light signal (LuxAB) α LuxI

Condition tables: describe the state of a species in the next step given the current states of all relevant species.

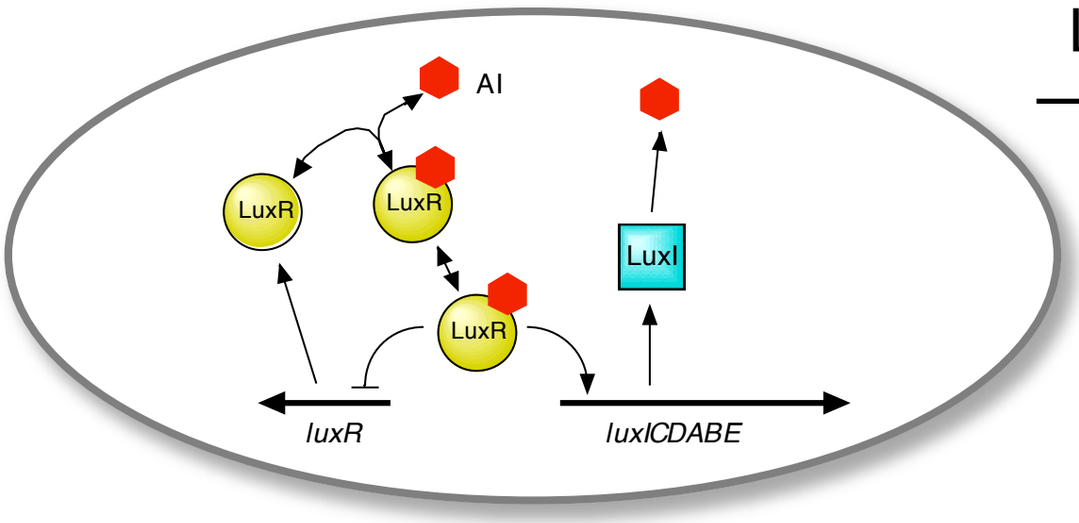
LuxI	LuxR:AI:Genome
0	0
1	1

How does LuxI depend on LuxR:AI:Genome?

LuxR:AI:Genome	LuxR:AI
0	0
1	1

How does LuxR:AI:Genome depend on LuxR:AI?

Condition Tables for QS II



LuxR	LuxR	AI	LuxR:AI:Genome	
1	0	0	0	When LuxR:AI:Genome is empty, LuxR is produced in next step
1	1	0	0	
1	0	1	0	
1	1	1	0	
0	0	0	1	
1	1	0	1	Comment: LuxR present, no AI available
0	0	1	1	
0	1	1	1	LuxR present, binds AI in next step, no LuxR is produced because LuxR:AI:Genome inhibits LuxR production

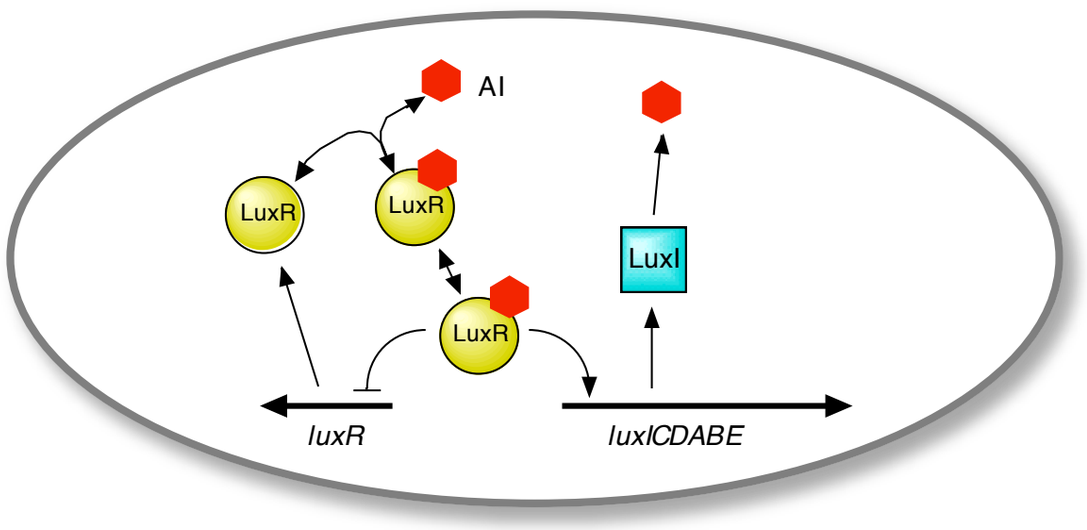
LuxR:AI	LuxR	AI	LuxR:AI:Genome
0	0	0	0
0	1	0	0
0	0	1	0
1	1	1	0
0	0	0	1
0	1	0	1
0	0	1	1
1	1	1	1

→

LuxR:AI	LuxR	AI	LuxR:AI:Genome
0	x	x	x
1	1	1	x

Note: no dissociation
(LuxR:AI:Genome → LuxR:AI + Genome)
only degradation of AI in this model
LuxR:AI:Genome → LuxR + Genome

Condition tables for QS III



AI	LuxR	AI	LuxI
0	0	0	0
0	1	0	0
1	0	1	0
0	1	1	0
1	0	0	1
1	1	0	1
1	0	1	1
1	1	1	1

→

AI	LuxR	AI	LuxI
1	x	x	1
0	x	0	0
1	0	1	0
0	1	1	0

Scanning for Attractors

States of *V. fischeri* QS system mapped onto integers

{LuxR (LR), LuxR:AI (RA), AI, LuxR:AI:Genome (RAG), LuxI (LI)}

= {1, 2, 4, 8, 16} - current state can be interpreted as binary number!

For each **attractor**:

- periodic orbit and its length (period)
- basin of attraction and its relative size (32 states in total)

→ how likely will the system end in each of the attractors?

Attractor 1: orbit: 1 → period 1

states: 0, 1 → size 2, 2/32 = 6.25 %

start from state 0:

#	LR	RA	AI	RAG	LI	-	state
0	.	.	.	.	.	-	0
1	X	.	.	.	.	-	1
2	X	.	.	.	.	-	1

<= attractor

Scanning for Attractors II

Attractor 2: orbit: 3, 9, 17, 5 → period 4

states: 2, 3, 5, 8, 9, 16, 17 → size 7, 21.9 %

start from state 8:

#	LR	RA	AI	RAG	LI	-	state
0	.	.	.	X	.	-	8
1	.	.	.	.	X	-	16
2	X	.	X	.	.	-	5
3	X	X	.	.	.	-	3
4	X	.	.	X	.	-	9
5	X	.	.	.	X	-	17
6	X	.	X	.	.	-	5

attractor

averaged occupancies in this periodic orbit:

LR	RA	AI	RAG	LI
4/4 = 1	1/4 = 0.25	1/4 = 0.25	1/4 = 0.25	1/4 = 0.25

Attractors III

Attractor 3: period 4, basin of 16 states → 50 %

#	LR	RA	AI	RAG	LI	– state0
	.	X	X	.	.	- 61
	.	X	X	X	.	- 142
	.	.	X	X	X	- 283
	.	.	X	.	X	- 20

Attractor 4: period 4, basin of 4 states → 12.5 %

#	LR	RA	AI	RAG	LI	- state0
	X	X	X	.	.	- 71
	X	X	.	X	.	- 112
	X	.	.	X	X	- 253
	X	.	X	.	X	- 21

Attractor 5: period 2, basin of 3 states → 9.4 %

#	LR	RA	AI	RAG	LI	- state0
	X	.	X	X	.	- 131
	.	X	.		X	- 18

Classifying the Attractors

→ Interpret the system's behavior from the properties of the attractors

Attractor	period	basin size	<LuxR>	<LuxR:AI>	<AI>	<LuxR:AI:Gen>	<LuxI>
1	1	6.25 % (2)	1	0	0	0	0
2	4	21.9% (7)	1	0.25	0.25	0.25	0.25
3	4	50 % (16)	0	0.5	1	0.5	0.5
4	4	12.5 % (4)	1	0.5	0.5	0.5	0.5
5	2	9.4% (3)	0.5	0.5	0.5	0.5	0.5

Three **regimes**:

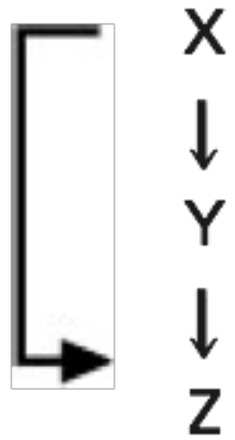
dark: LuxI = 0 — free LuxR, no AI

intermediate: LuxI = 0.25 — free LuxR + little AI

bright: LuxI = 0.5 — little free LuxR (0.24) + much AI (0.85)

The Feed-Forward-Loop

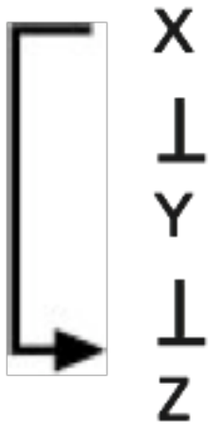
External signal determines state of X
→ response Z for short and long signals X



condition tables:

Y	X
0	0
1	1

Z	X	Y
0	0	0
0	0	1
0	1	0
1	1	1



Y	X
1	0
0	1

Z	X	Y
0	0	0
0	0	1
1	1	0
0	1	1

Signal propagation
Left column: external signal

Response to signal X(t)

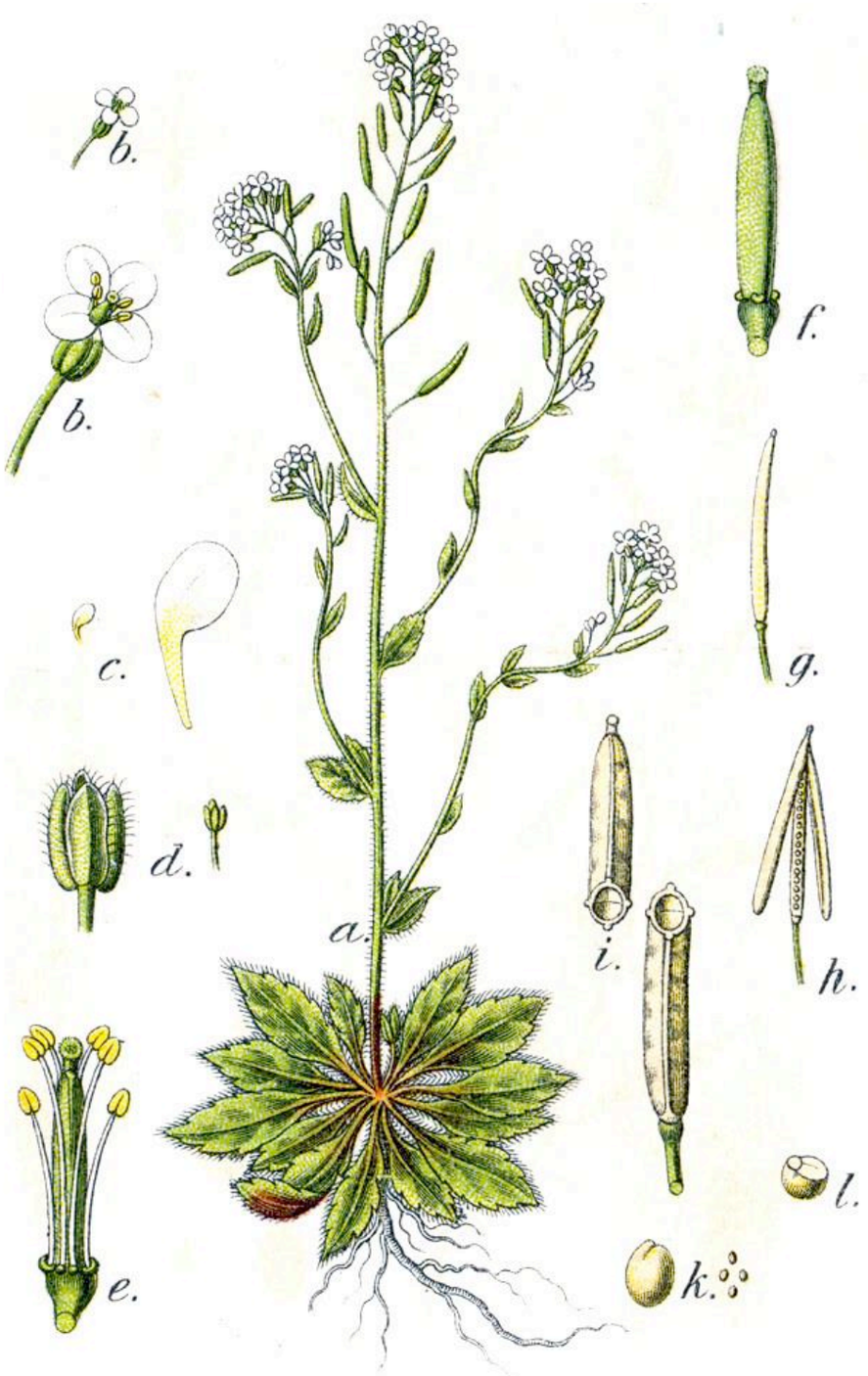
X	Y	Z
0	0	0
1	0	0
0	1	0
0	0	0
1	0	0
1	1	0
1	1	**1**
0	1	**1**
0	0	**0**
0	0	0

Short Signal

Long signal

X	Y	Z
0	1	0
1	1	0
0	0	0
0	1	0
1	1	0
1	0	0
1	0	**1**
0	0	**1**
0	1	**1**
0	1	0

The *A. thaliana* Flowering Network



Model organism in genomics:

- small, convenient to grow
- completely sequenced (2000): 125 Mbp
- can be easily mutated

also see: Arabidopsis Information Resource (TAIR)@ www.arabidopsis.org/

images from wikimedia

Dynamics of the Genetic Regulatory Network for *Arabidopsis thaliana* Flower Morphogenesis

LUIS MENDOZA AND ELENA R. ALVAREZ-BUYLLA*

Lab. de Genética Molecular y Evolución, Instituto de Ecología, UNAM, Ap. Postal 70-275, *Coyoacán, D.F. CP*04510, *Mexico*

(*Received on* 26 *September* 1997, *Accepted in revised form on* 3 *March* 1998)

We present a network model and its dynamic analysis for the regulatory relationships among 11 genes that participate in *Arabidopsis thaliana* flower morphogenesis. The topology of the network and the relative strengths of interactions among these genes were based from published genetic and molecular data, mainly relying on mRNA expression patterns under wild type and mutant backgrounds. The network model is made of binary elements and we used a particular dynamic implementation for the network that we call semi-synchronic. Using this method the network reaches six attractors; four of them correspond to observed patterns of gene expression found in the floral organs of *Arabidopsis* (sepals, petals, stamens and carpels) as predicted by the ABC model of flower morphogenesis. The fifth state corresponds to cells that are not competent to flowering, and the sixth attractor predicted by the model is never found in wild-type plants, but it could be induced experimentally. We discuss the biological implications and the potential use of this network modeling approach to integrate functional data of regulatory genes of plant development.

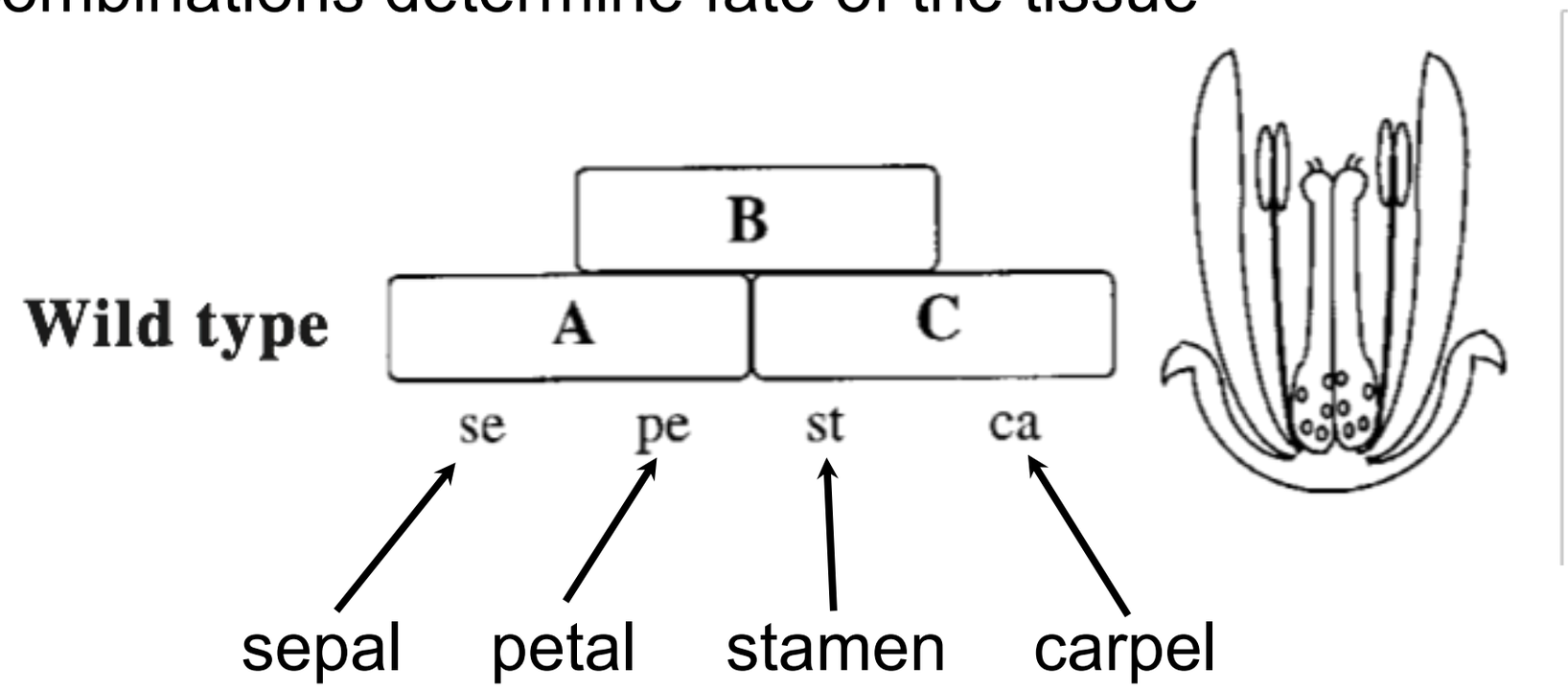
J. theor Biol. **193** (1998) 307

The ABC Model

Coen, Meyerowitz (1991):

three different activities A, B, and C, active in two adjacent whorls,
mutual inhibition of A and C

→ combinations determine fate of the tissue



Related genes:

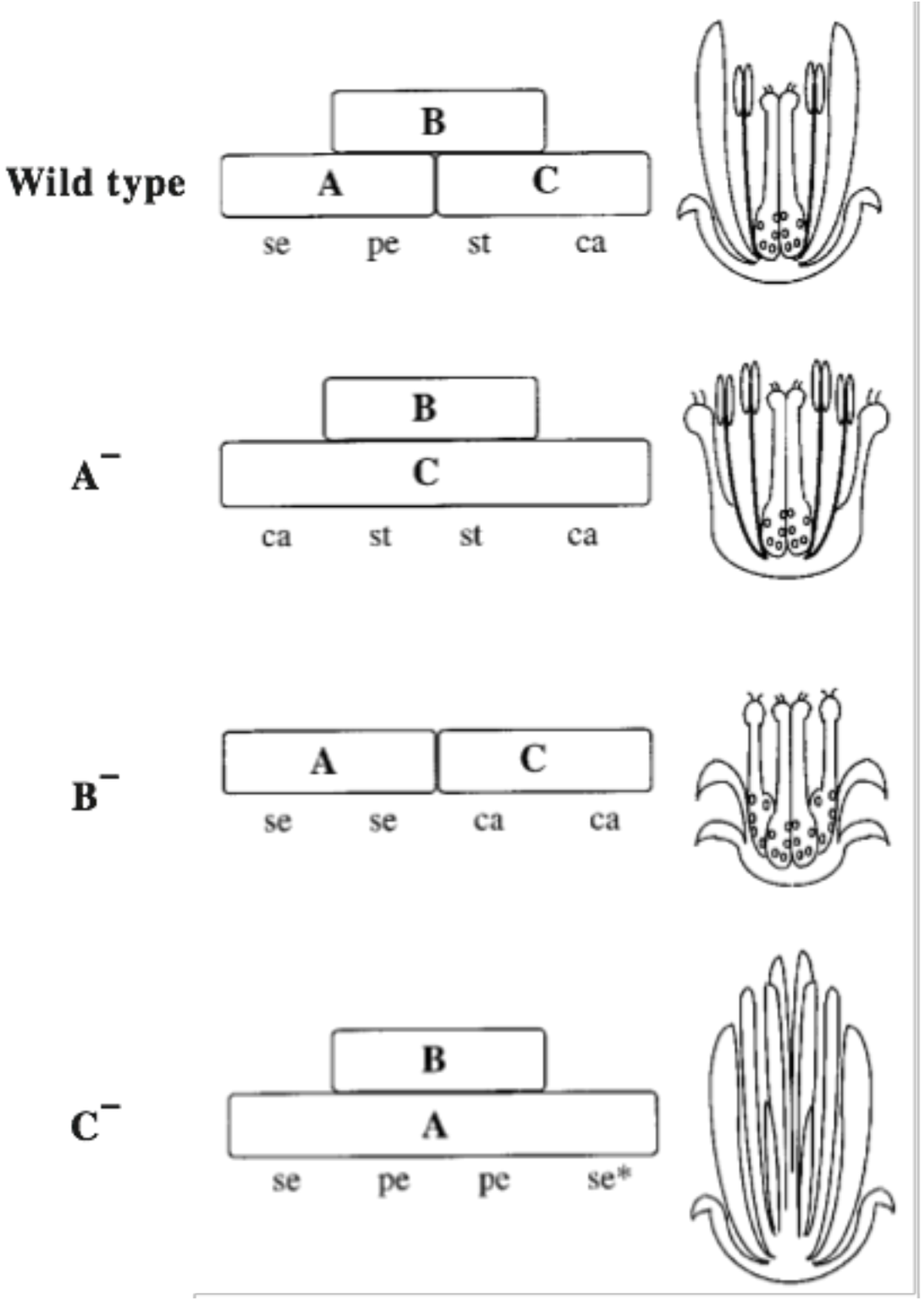
A:
APETALA1 (*AP1*)

B:
APETALA3 (*AP3*),
PISTILATA (*PI*)

C:
AGAMOUS (*AG*)

Mendoza, Alvarez-Buylla, *J. theor Biol.* **193** (1998) 307

ABC Mutants

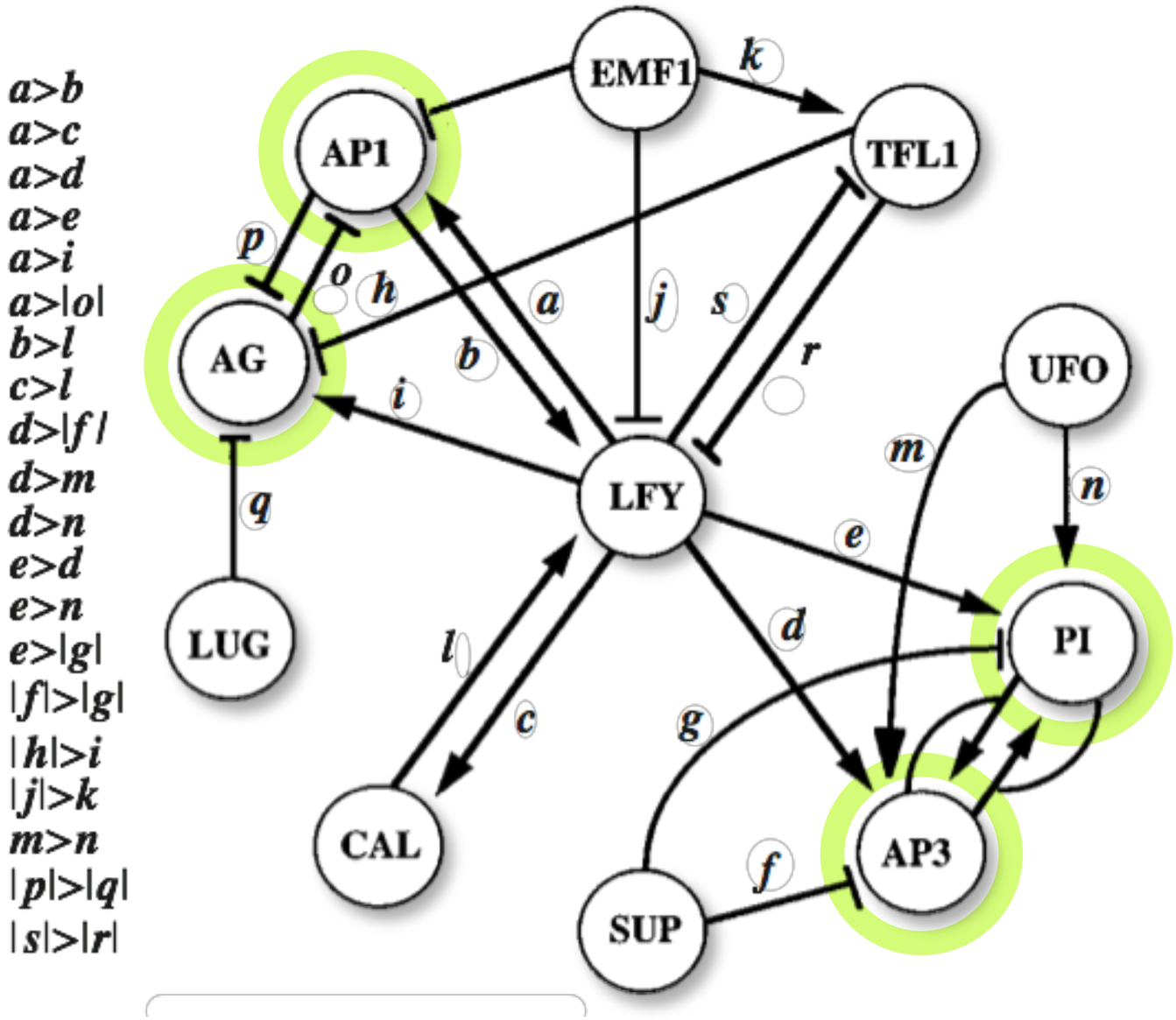


If any of the 3 functions (activities) is missing, the flowers have different tissue combinations.

se = sepals,
pe = petals,
st = stamens,
ca = carpels,
se* = se, pe, pe

The Network Model

11 genes (including the four ABC genes)



inequalities denote the relative weights of the interactions

Mendoza, Alvarez-Buylla, *J. theor Biol.* **193** (1998) 307

Model Implementation

Here: Boolean model with **weighted** interactions

Propagate state vector $\boldsymbol{x} = \{x_1, x_2, \ldots, x_{11}\}$ by:

$$x_i(t+1) = \mathbf{H}\left(\sum_{j=1}^{N} w_{ij} x_j(t) - \theta_i\right)$$

Heavyside step function: $\mathbf{H}(x) = \begin{cases} 1 & \text{if} \quad x > 0 \\ 0 & \text{if} \quad x \leqslant 0 \end{cases}$

Weights w_{ij} and thresholds θ_i are not known exactly
→ choose integers for simplicity
→ positive for activation, negative for inhibition

Mendoza, Alvarez-Buylla, *J. theor Biol.* **193** (1998) 307

The Numbers

*EMF*1, *TFL*1, *LFY*, *AP*1, *CAL*, *LUG*, *UFO*, *BFU*,*AG*, *AP*3, *PI* and *SUP*.

$$\mathbf{W} = \begin{bmatrix}
0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\
1 & 0 & -2 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\
-2 & -1 & 0 & 2 & 1 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\
-1 & 0 & 5 & 0 & 0 & 0 & 0 & 0 & -1 & 0 & 0 & 0 \\
0 & 0 & 2 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\
0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\
0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\
0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 1 & 1 & 0 \\
0 & -2 & 1 & -2 & 0 & -1 & 0 & 0 & 0 & 0 & 0 & 0 \\
0 & 0 & 3 & 0 & 0 & 0 & 2 & 1 & 0 & 0 & 0 & -2 \\
0 & 0 & 4 & 0 & 0 & 0 & 1 & 1 & 0 & 0 & 0 & -1 \\
0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0
\end{bmatrix} \qquad \boldsymbol{\theta} = \begin{bmatrix}
0 \\ 0 \\ 3 \\ -1 \\ 1 \\ 0 \\ 0 \\ 1 \\ -1 \\ 0 \\ 0 \\ 0
\end{bmatrix}.$$

Mendoza, Alvarez-Buylla, *J. theor Biol.* **193** (1998) 307

Synchronous vs. Asynchronous

Synchronous propagation (Kauffman (1969)):

→ update all species **simultaneously**

→ biological problem: do all genes respond at exactly the same time?

Asynchronous propagation (Thomas (1991)):

→ update one species after the other **in chosen order**

→ order of update may influence dynamic gene activation patterns

Semi-synchronic propagation (Mendoza (1998)):

→ split genes in groups:

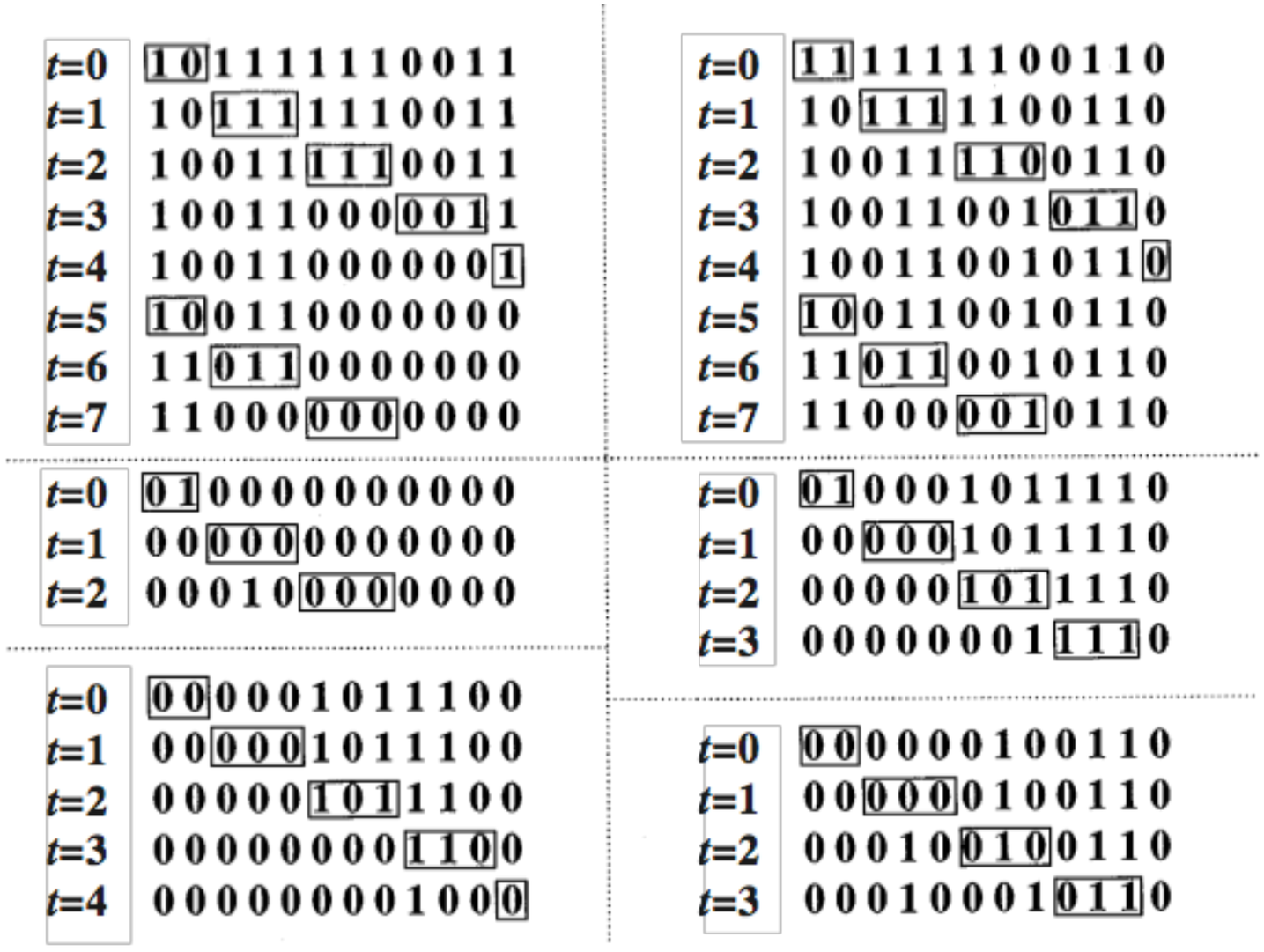
→ synchronous within group, one group after the other

→ base order of groups upon experimental data (but it's still a "choice")

*EMF*1, *TFL*1 → *LFY, AP*1, *CAL* → *LUG, UFO, BFU* → *AG, AP*3, *PI* → *SUP*

Mendoza, Alvarez-Buylla, *J. theor Biol.* **193** (1998) 307

Some Example Patterns



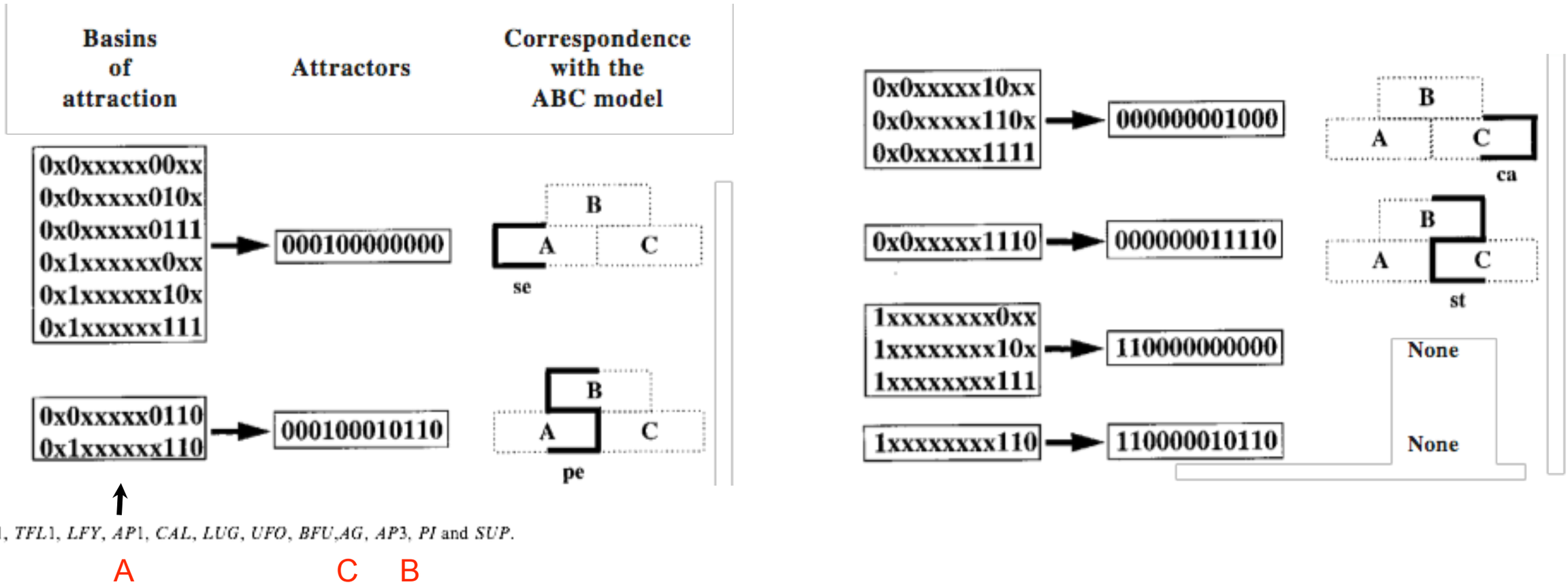
Exhaustive search: start from **all** 2^{12} = 4096 possible **initial** states,

run for t = 200 steps

→ **six stationary** patterns (attractors of size 1)

Mendoza, Alvarez-Buylla, *J. theor Biol.* **193** (1998) 307

The Attractors



From gene **activation patterns** in the attractors:

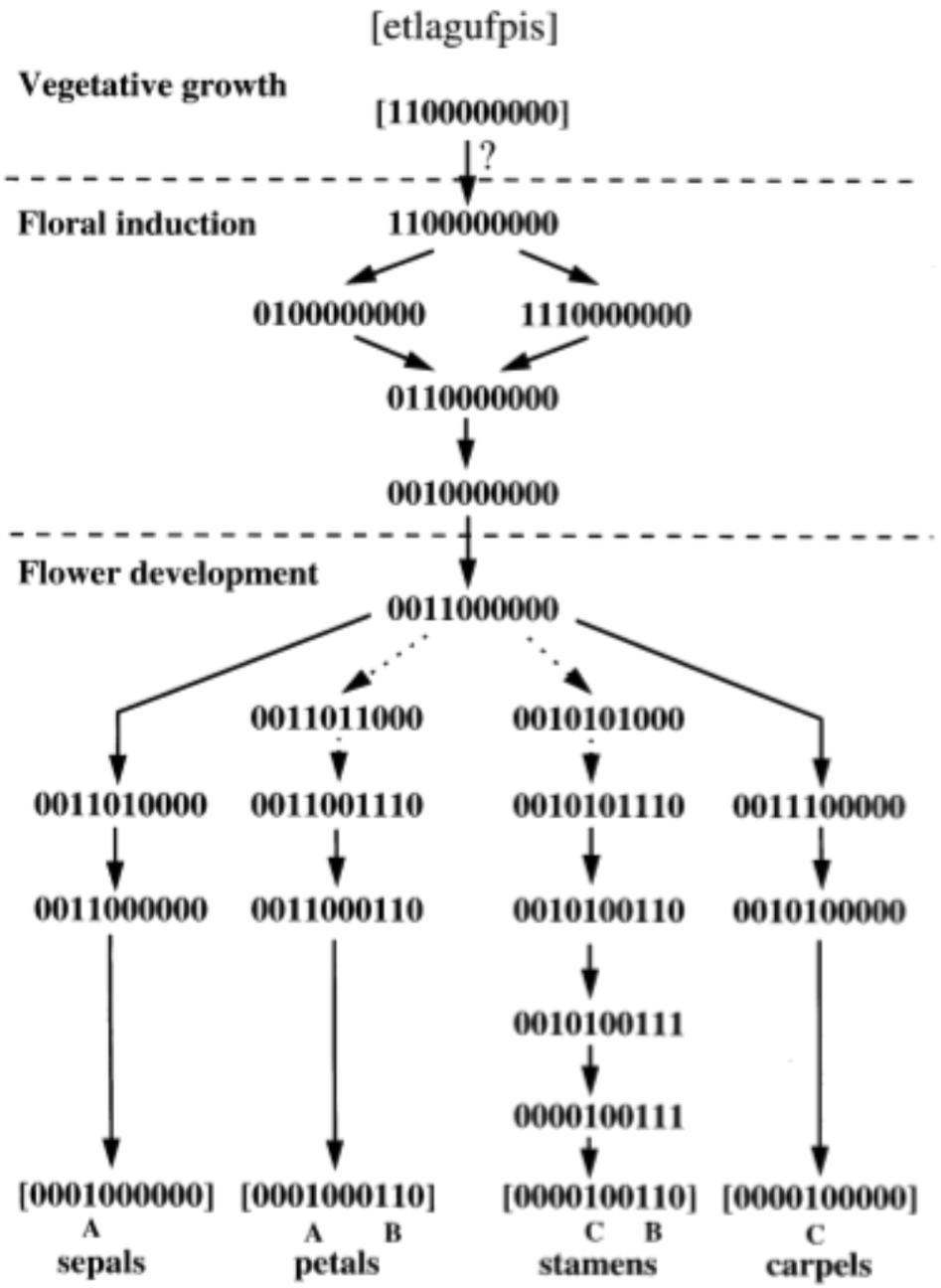
→ identify the **four floral** tissue **types** of the ABC model

→ one attractor with floral **inhibitors** *EMF*1, *TFL*1
(characteristic for cells that are not part of the flowers)

→ one yet **unidentified** state

Mendoza, Alvarez-Buylla, *J. theor Biol.* **193** (1998) 307

Possible Pathways of flower development



Note: the model does not include temporal and spatial information required to predict where and when which genes are activated or repressed ("signals")

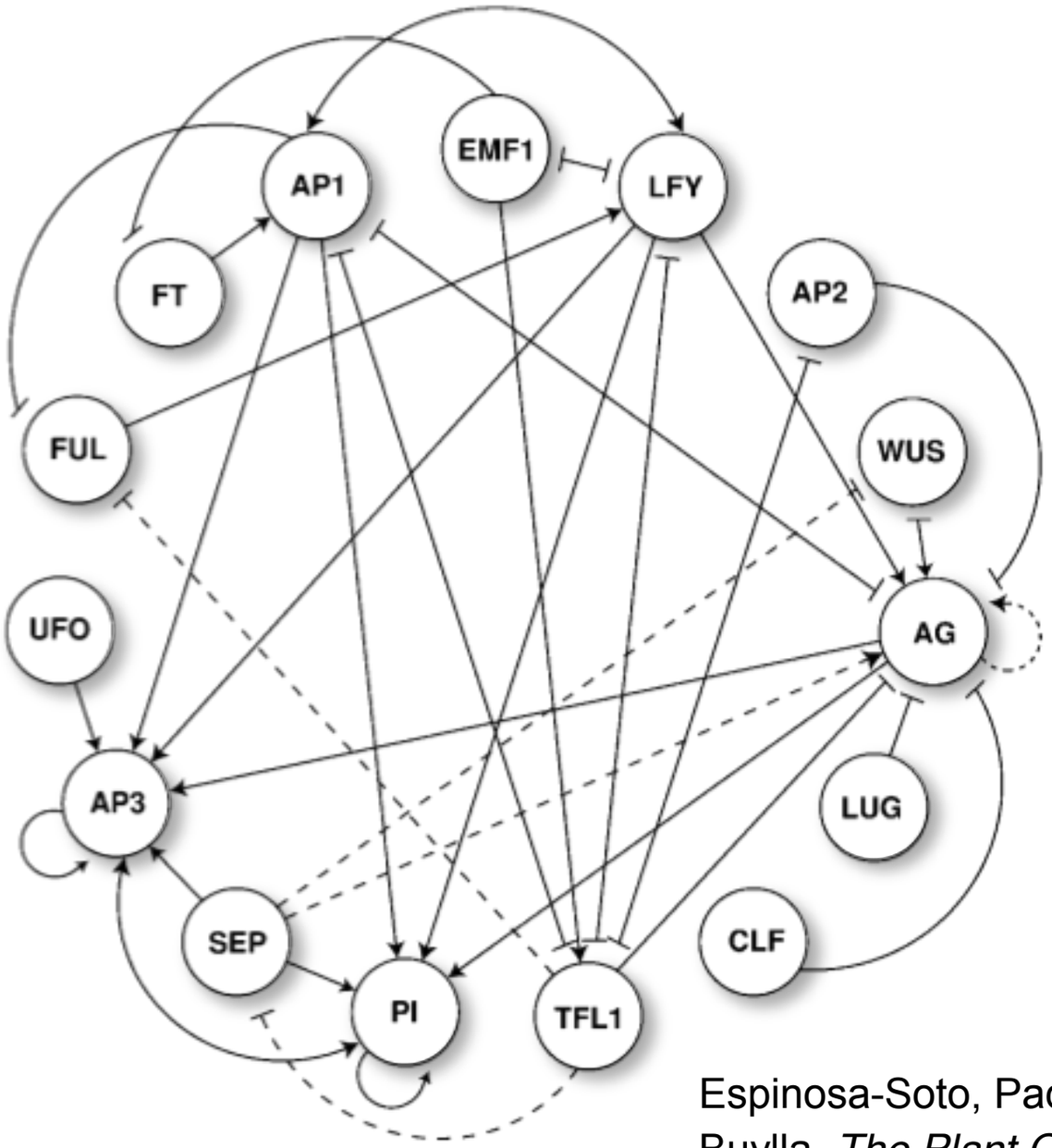
→ these pathways are a "proposal" only

Mendoza et al, *Bioinformatics* **15** (1999) 593

Sophistication of Networks

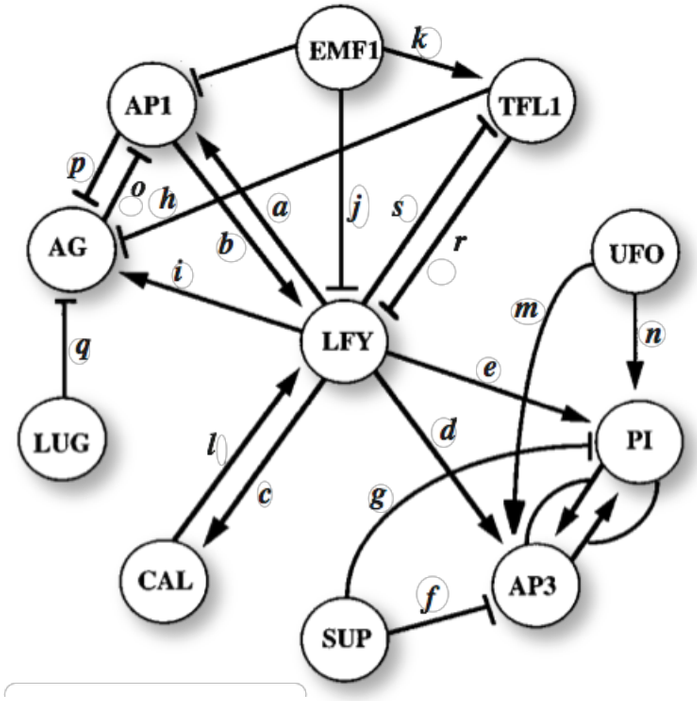
A few years later:

additional genes and predicted interactions (- - -) more states



2004

1998



Mendoza, Alvarez-Buylla, *J. theor Biol.* **193** (1998) 307

Espinosa-Soto, Padilla-Longoria, Alvarez-Buylla, *The Plant Cell* **16** (2004) 2923

What is it Worth?

Generally: → quality of the **results** depends on the quality of the **model**

→ quality of the model depends on the quality of the **assumptions**

Assumptions for the Boolean network description:

(• subset of the species considered → reduced system state space)

- only discrete density levels
- conditional yes–no causality

→ dynamic balances lost,
reduced to oscillations
→ no continuous processes

- discretized propagation steps

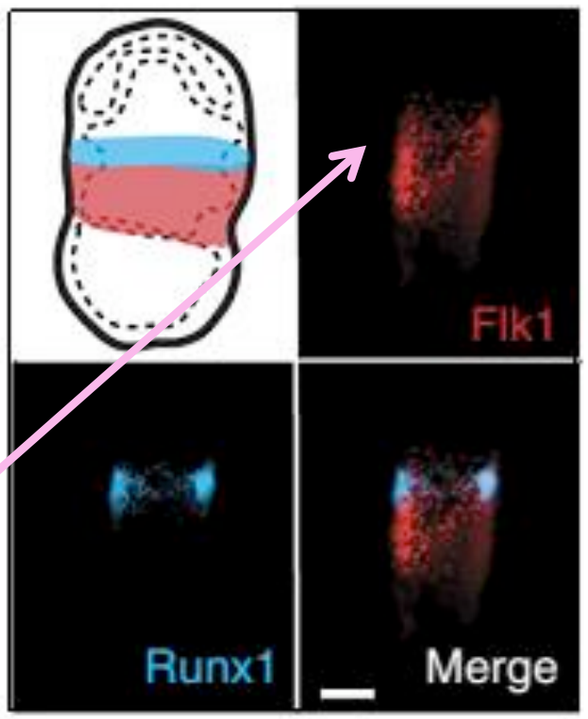
→ timing of concurrent paths?

"You get what you pay for"

Understand Blood development (hemato-poeisis) with the help of Boolean Networks

Blood has long served as a model to study organ development owing to the **accessibility** of blood cells and the availability of markers for specific cell populations.

Blood development is initiated at **gastrulation** from multipotent Flk1$^+$ mesodermal cells
(Flk1$^+$ is a marker gene for this developmental stage.)



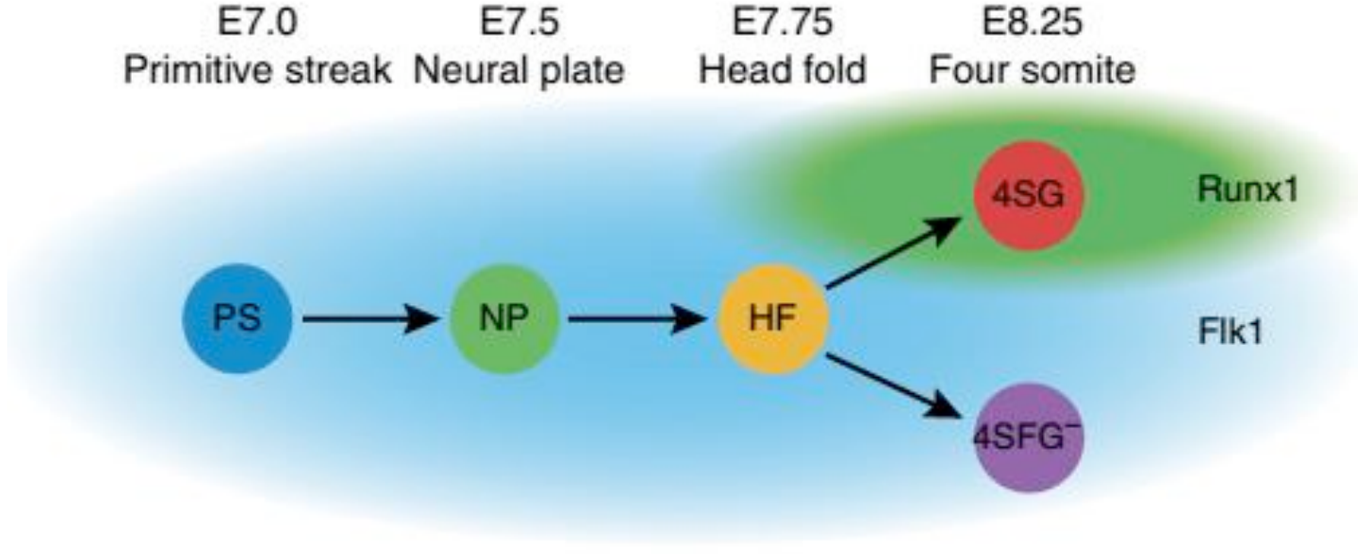
Flk1 and Runx1 staining in E7.5 mesoderm and blood band, respectively

These cells initially have the potential to form blood, endothelium and smooth muscle cells.

Blood development represents one of the **earliest stages** of **organogenesis**, as the production of primitive erythrocytes is required to support the growing embryo.

Moignard et al., Nature Biotech. 33, 269 (2015)

Early stages of hematopoesis



The first wave of primitive hematopoiesis originates from Flk1$^+$ mesoderm, with all hematopoietic potential in the mouse contained within the Flk1$^+$ population from E7.0 onwards.

Authors of this study flow sorted single Flk1$^+$ cells at E7.0 (primitive streak, PS), E7.5 (neural plate, NP) and E7.75 (head fold, HF) stages.

They subdivided E8.25 cells into putative blood and endothelial populations by isolating GFP$^+$ cells (four somite, 4SG) and Flk1$^+$GFP$^-$ cells (4SFG−), respectively

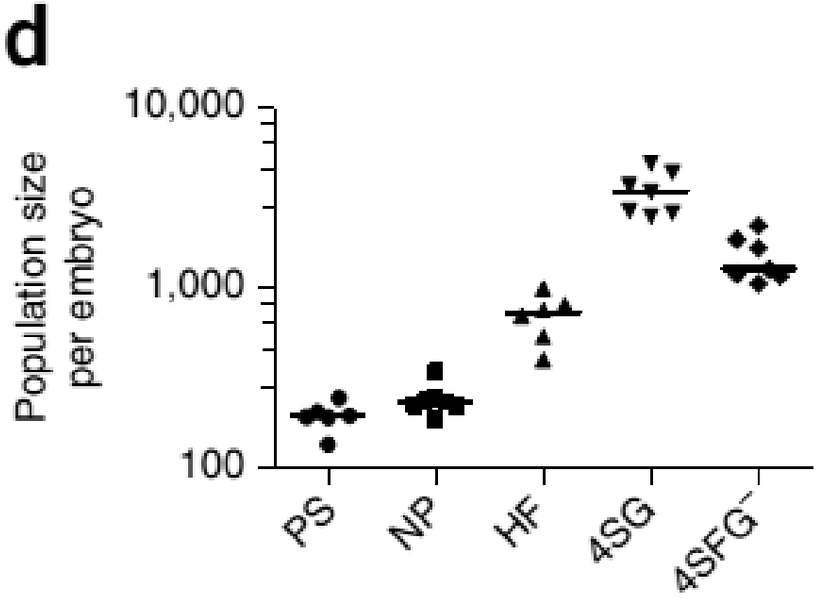
Moignard et al.,
Nature Biotech.
33, 269 (2015)

Studied cells

Cells were sorted from multiple embryos at each time point, with 3,934 cells going on to subsequent analysis.

Cell type	Number of embryos	Cells sorted	Cells retained	Percentage retained
PS	12	725	624	86.1
NP	9	637	552	86.7
HF	8	1,184	1,005	84.9
4SG	3	1,085	983	90.6
4SFG$^-$	4	858	770	89.7
Total	36	4,489	3,934	87.6

Total cell numbers and numbers of cells of appropriate phenotypes present in each embryo were estimated from fluorescence-activated cell sorting (FACS) data.



Number of cells grows as embryo development progresses

Moignard et al.,
Nature Biotech.
33, 269 (2015)

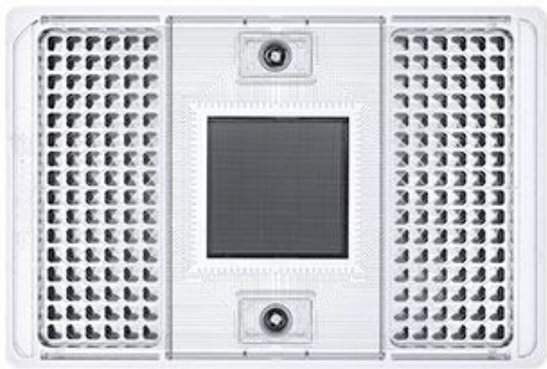
Assay gene expression in single cells

Cell type	Number of embryos	Cells sorted	Cells retained	Percentage retained
PS	12	725	624	86.1
NP	9	637	552	86.7
HF	8	1,184	1,005	84.9
4SG	3	1,085	983	90.6
4SFG$^-$	4	858	770	89.7
Total	36	4,489	3,934	87.6

Discard cells that did not express all 4 house-keeping genes, or for which their expression was more than 3 standard deviations from the mean.

Gene expression in single cells assayed with PCR for:

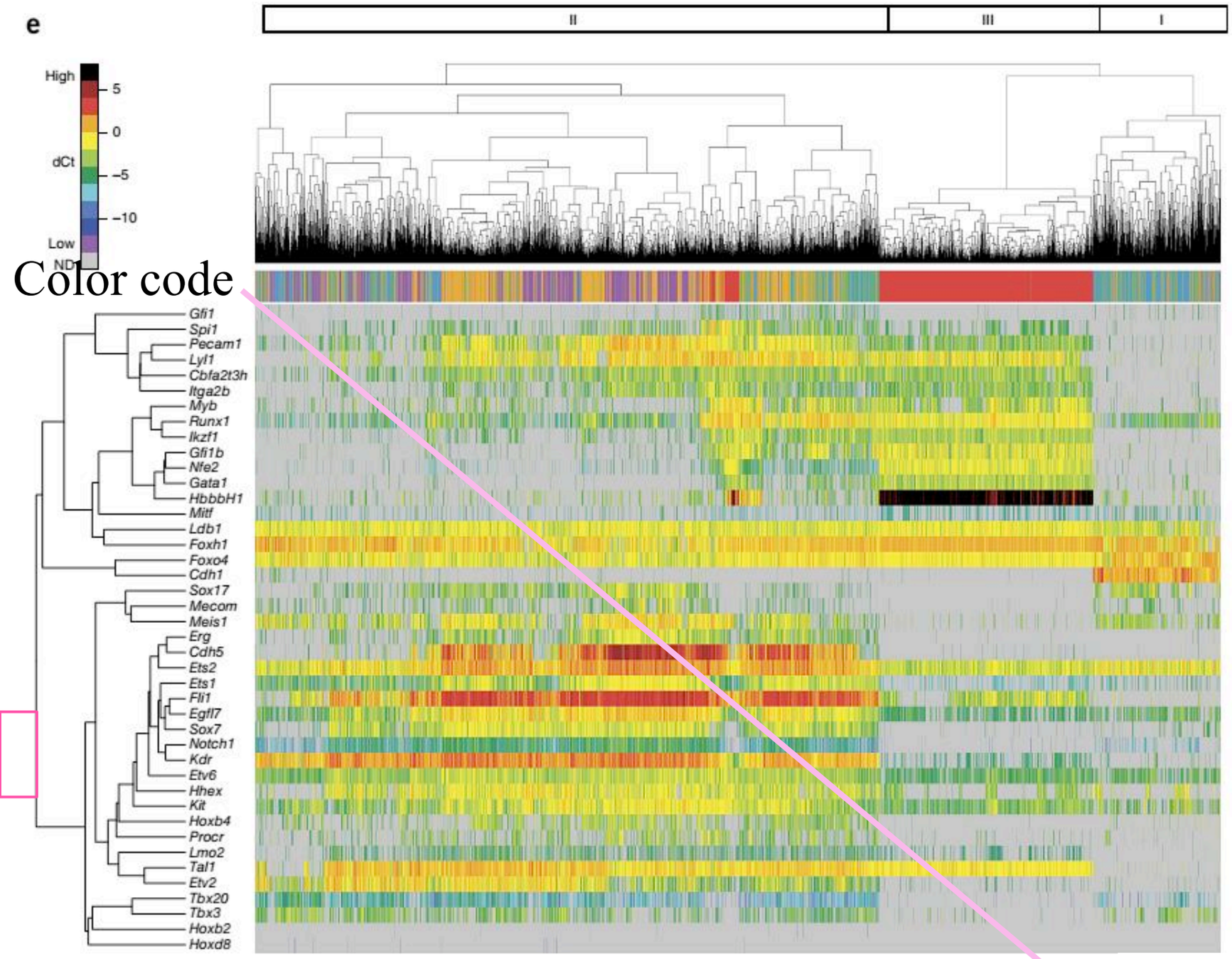
- 33 transcription factors known to be involved in endothelial and hematopoietic development
- 9 marker genes (needed for FACS-sorting)
- 4 house-keeping genes (needed for quality checks and normalization)



www.fluidigm.com

Moignard et al., Nature Biotech. 33, 269 (2015)

Hierarchical clustering of gene expression data



← Single cells →

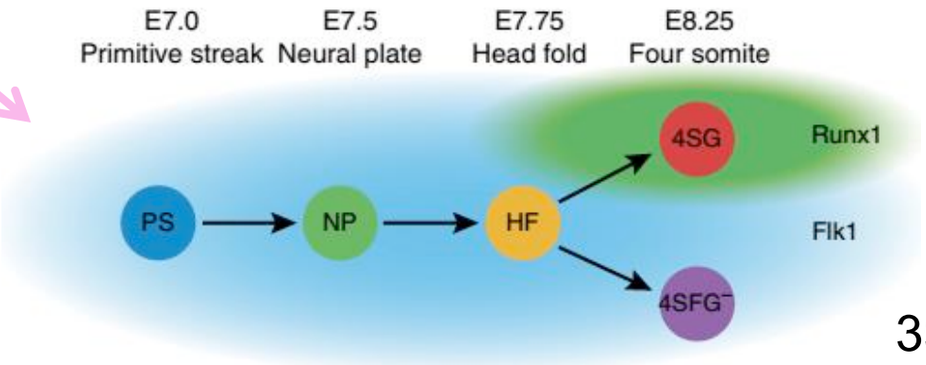
3 main clusters:

Cluster I (right side) contains mostly PS and NP cells

Cluster III contains exclusively 4SG cells

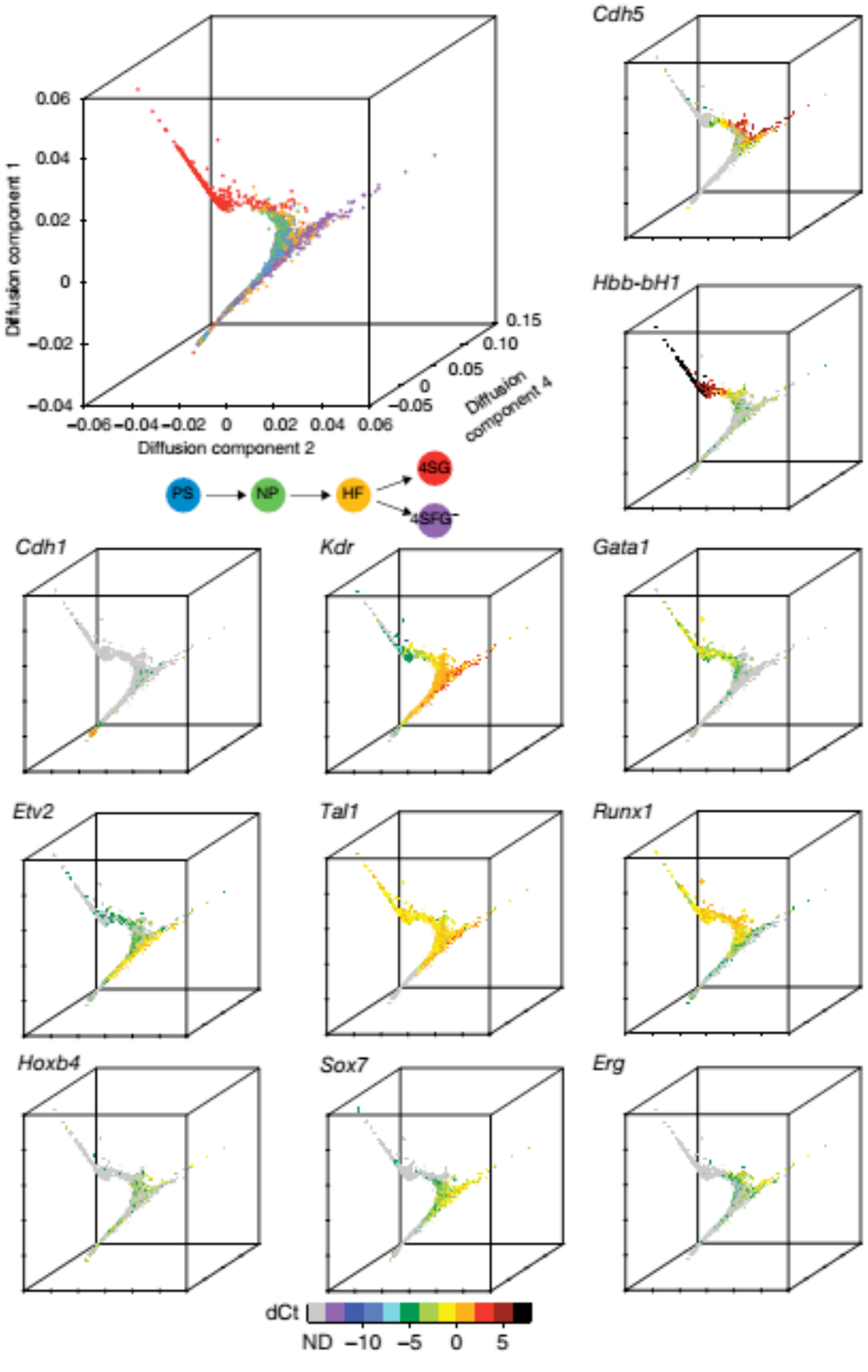
Cluster II is mixed (NF, 4SFG$^-$, …)

→ Cell differentiation progresses asynchronously



Moignard et al., Nature Biotech. 33, 269 (2015)

Dimensionality reduction: diffusion maps



Similarity of expression in cells *i* and *j* :

$$P(i,j) = \frac{1}{Z_i} \exp\left(\frac{-\left(x_i - x_j\right)^2}{\varepsilon} \right)$$

P(i,j) is normalized so that $\sum_{i=1} P(i,j) = 1$

The cells are organized in 2D or 3D such that the Euclidean distance between the cells corresponds to the diffusion metric *P*(*i*,*j*)

.

The quantity *P*(*i*,*j*) can then be interpreted as the transition probability of a diffusion process between cells.

Axes: eigenvectors of matrix P with largest eigenvalues.

Moignard et al., Nature Biotech. 33, 269 (2015) V10 –

Who regulates hematopoiesis?
Design Boolean Network

33 transcription factors
3,934 cells
= 129,822 RTqPCRs

Binary discretization →

Possible binary states =	$2^{33} \approx 8{,}589 \times 10^6$
Measured binary states =	3,934
Observed unique binary states =	3,070
Largest connected component =	1,448

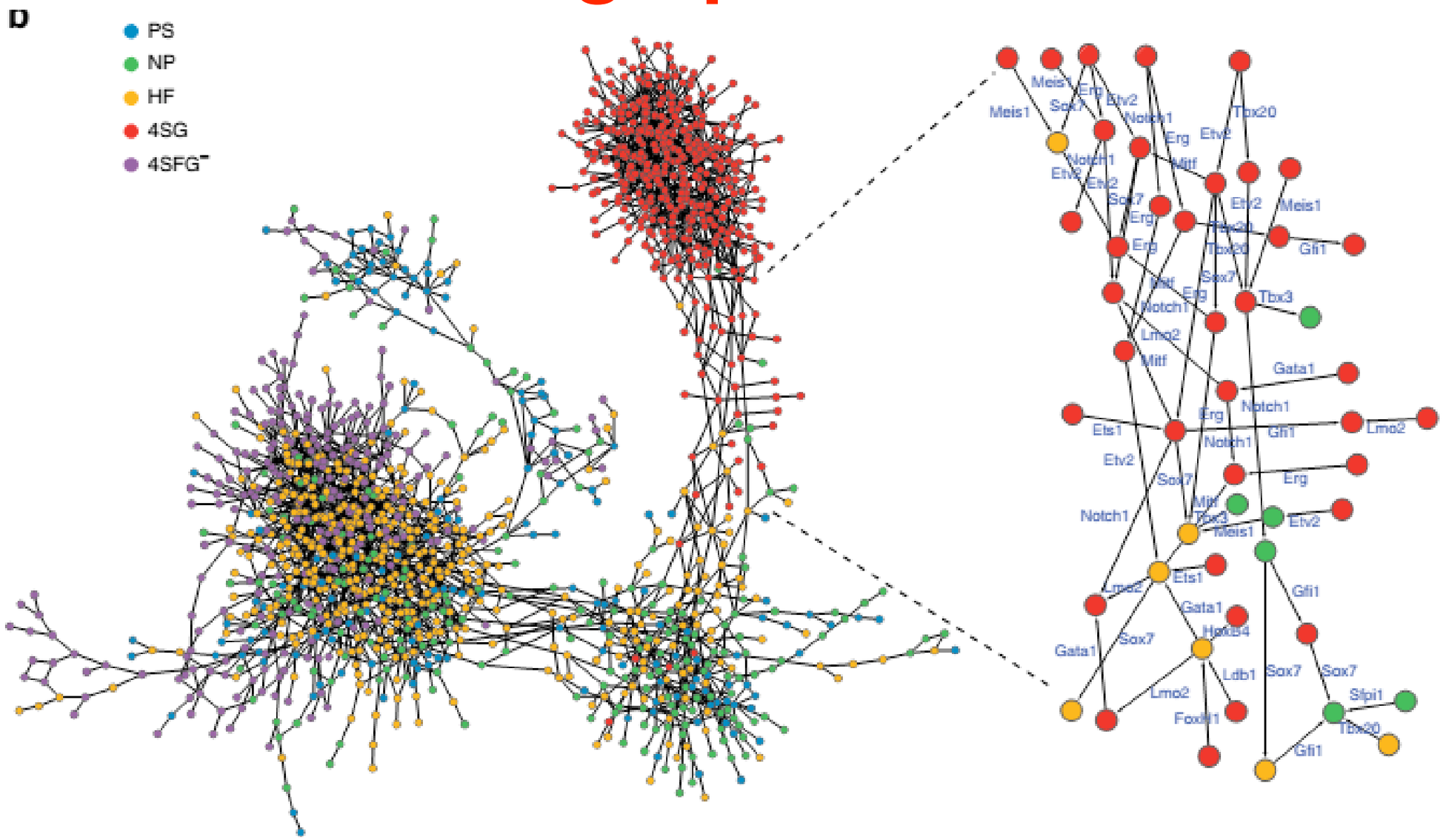
Determine suitable expression thresholds for each gene to categorize its expression levels into binary on / off states.

Note that less than 0.1% of the possible states has been observed.

Moignard et al.,
Nature Biotech.
33, 269 (2015)

State graph

Moignard et al.
Nature Biotech
33, 269 (2015)



State graph (largest connected component) of 1448 states reaching all 5 stages.

Edges connect all states that differ in the on/off levels of a single gene.

Automatic derivation of rules for Boolean Network

We are given:

- a set of variables V, corresponding to **genes**,

- an undirected graph $G = (N,E)$
  where each node $n \in N$ is labeled with a **state** $s: V \rightarrow \{0,1\}$, and
  each edge $\{s_1,s_2\} \in E$ is labeled with the single variable
  that changes between state s_1 and s_2.

We are also given a designated set $I \subseteq N$ of **initial vertices**
and a designated set $F \subseteq N$ of **final vertices**,
along with a **threshold** t_i for each variable $v_i \in V$.

Moignard et al.,
Nature Biotech.
33, 269 (2015)

Optimality criteria for rules

Our synthesis method searches for an orientation of G, along with an update function u_i:$\{0,1\}^n \rightarrow \{0,1\}$ for each variable $v_i \in V$, such that the following conditions hold:

1. For each edge (s_1,s_2) labeled with variable v_i in the orientated graph, the update function for v_i takes state s_1 to state s_2: $u_i(s_1) = s_2(i)$.

2. The number of states in which no transitions induced by the update functions are missing is maximized.

3. Every final vertex $f \in F$ is reachable from some initial vertex $i \in I$ by a directed path in the orientated graph.

Moignard et al.,
Nature Biotech.
33, 269 (2015)

Allowed complexity of the rules

We restrict the update function u_i to have the form:

$$f_1 \wedge \neg f_2$$

where f_j is a Boolean formula that has

and-nodes of in-degree two,

or-nodes of arbitrary in-degree, and

where f_1 has a maximum depth of N_i and f_2 has a maximum depth of M_i.

N_i and M_i are given as parameters to the method.

The search for edge orientations and associated Boolean update rules is encoded as a Boolean satisfiability (SAT) problem.

Moignard et al.,
Nature Biotech.
33, 269 (2015)

Generated rules for Boolean Network

Gene	Synthesised update functions	% Non-observed transitions disallowed (N_i)	Motifs present
Scl	*Fli1*	98	Yes
Etv2	*Notch1*	96	Yes
Fli1	*Etv2*	96	Yes
	Sox7	97	Yes
Lyl1	*Sox7*	92	Yes
Sox7	*Sox17* ∨ *HoxB4*	82	No (Sox missing)
Erg	(*HoxB4* ∧ *Lyl1*) ∨ *Sox17*	84	Yes
	(*HoxB4* ∧ *Tal1*) ∨ *Sox17*	83	Yes
Notch1	*Sox7*	94	Yes
Gata1	*Gfi1b* ∧ *Lmo2*	86	Yes
	Gfi1b ∧ *Hhex*	84	No (Hhex missing)
	Gfi1b ∧ *Ets1*	84	Yes
HoxB4	(*Lyl1* ∧ *Ets1*) ∧ ¬*Gata1*	65	Yes
	(*Lyl1* ∨ *Nfe2*) ∧ ¬*Gata1*	65	Yes
	(*Lyl1* ∨ *Ikaros*) ∧ ¬*Gata1*	65	No (Ikaros missing)
Sox17	*Lyl1* ∧ ¬*Gfi1b*	77	No (Gfi missing)
	(*Eto2* ∧ *Sox7*) ∧ ¬*Gfi1b*	76	No (Gfi missing)
	(*Eto2* ∧ *Tal1*) ∧ ¬*Gfi1b*	75	No (Gfi missing)
Ets1	*Notch1*	96	Yes
Gfi1	*Gata1* ∧ ¬*Sox17*	88	Yes
	Nfe2 ∧ ¬*Sox17*	88	Yes
Gfi1b	*Nfe2* ∧ *Myb*	87	Yes
	Pu.1 ∧ *Ikaros*	86	No (Ikaros missing)
	Pu.1 ∧ *Nfe2*	86	Yes
	Pu.1 ∧ *Myb*	86	Yes
Eto2	*Sox7*	93	No (Sox missing)
	Hhex	92	No (Hhex missing)
	Ets1 ∧ *Fli1*	94	No (Ets missing)
Hhex	*Sox7*	97	No (Sox missing)
	Notch1	93	No (Rbpj missing)
Ikaros	*Nfe2* ∨ *Gfi1b*	84	Yes
	Nfe2 ∨ *Gata1*	83	Yes
	Nfe2 ∨ *Gfi1*	82	Yes
Lmo2	*Sox7* ∨ *Gfi1*	79	Yes
	Sox7 ∨ *Erg*	79	Yes
	Sox7 ∨ *HoxB4*	77	Yes
Nfe2	*Ikaros*	78	Yes
Pu.1	*Gfi1* ∨ *Erg*	67	Yes
Myb	*HoxB4*	64	Yes

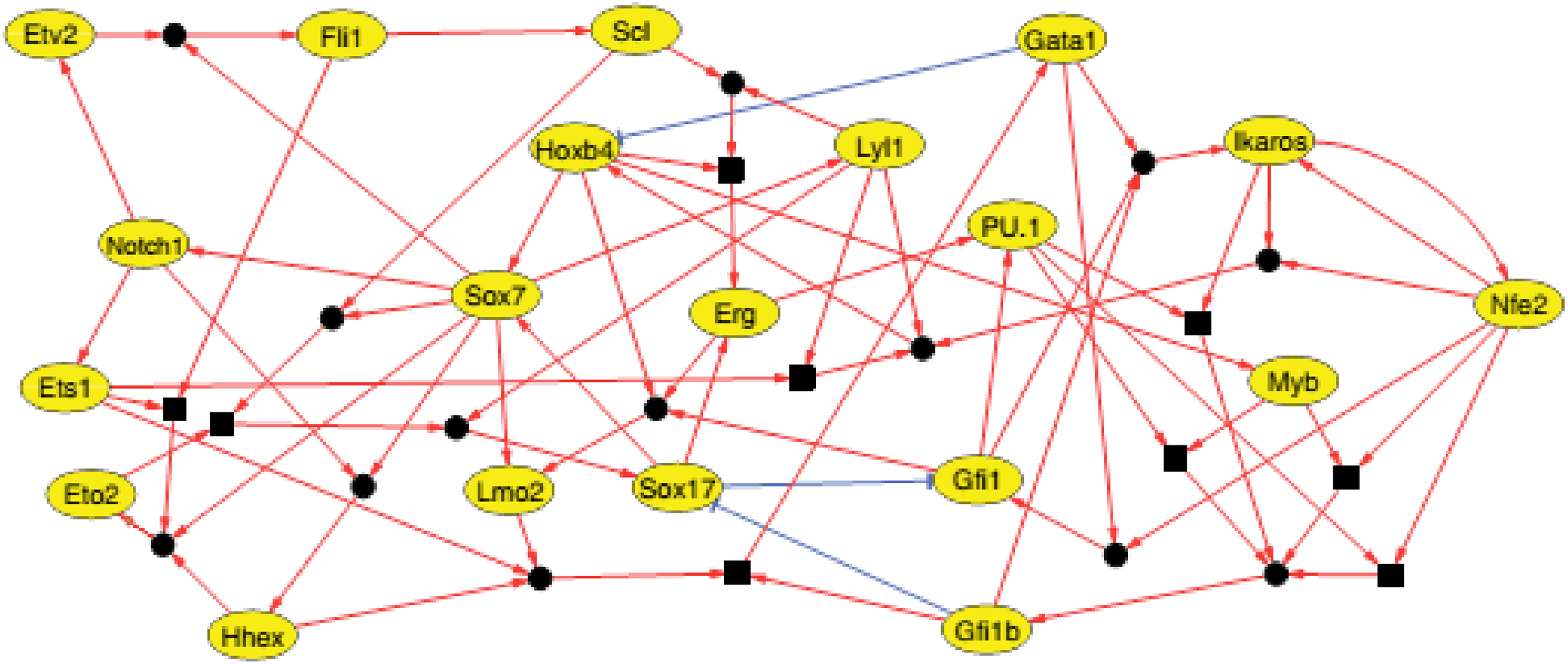
→ Additional validity check of the postulated rules:

check whether regulated genes contain TF-binding motifs in their promoters (right column).

This is the case for 70% of the rules.

Moignard et al.,
Nature Biotech.
33, 269 (2015)

Core network controlling hematopoiesis



Derived core network of 20 TFs.

Red edges: activation
Blue edges: repression

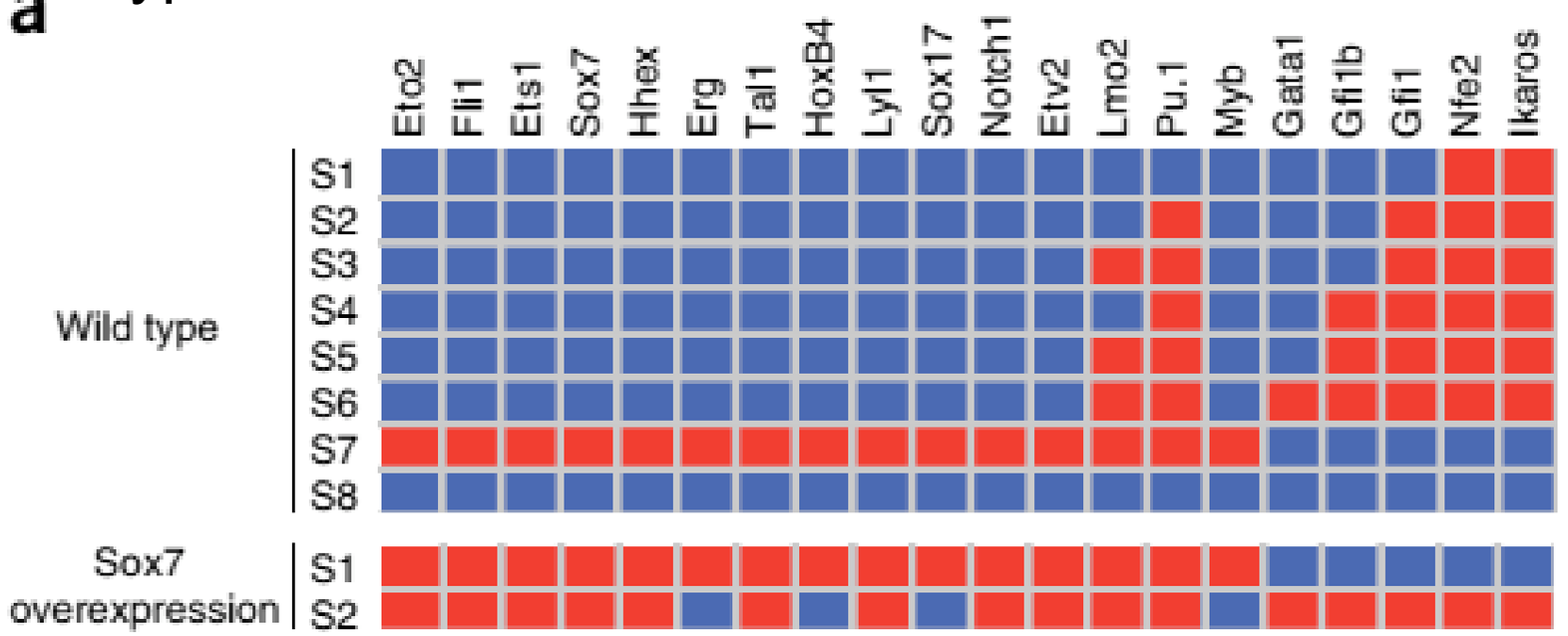
Moignard et al.,
Nature Biotech.
33, 269 (2015)

Predict effects of perturbations as validation

In silico perturbations predict key regulators of blood development.
Overexpression and knockout experiments were simulated for each TF and the ability of the network to reach wildtype or new stable states was assessed

Red indicates expressed;
blue indicates not expressed.

S2-S6: blood-like
S7: endothelial-like



Network stable states for wt and Sox7 overexpression.

Enforced expression of Sox7 (that is normally downregulated) stabilized the endothelial module and an inability to reach any of the blood-like states.

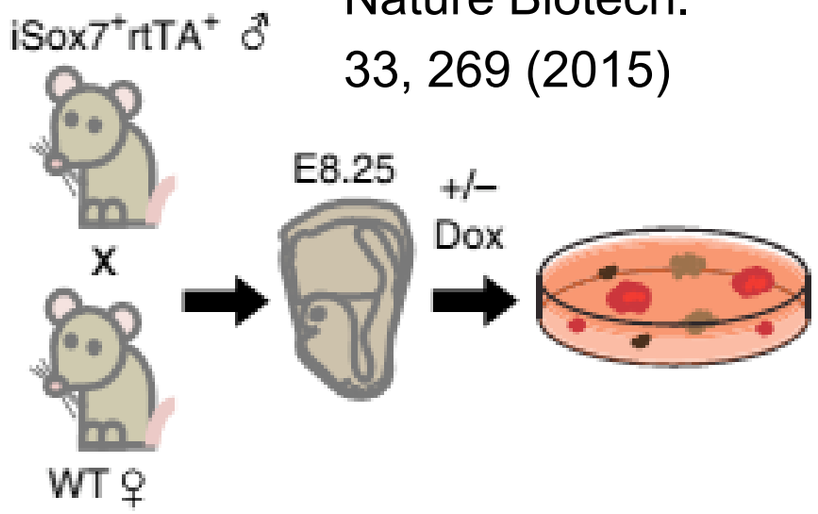
Sox7 is predicted to regulate more targets than any other TF, suggesting that perturbing its expression could have important downstream consequences

Moignard et al.,
Nature Biotech.
33, 269 (2015)

Control experiments

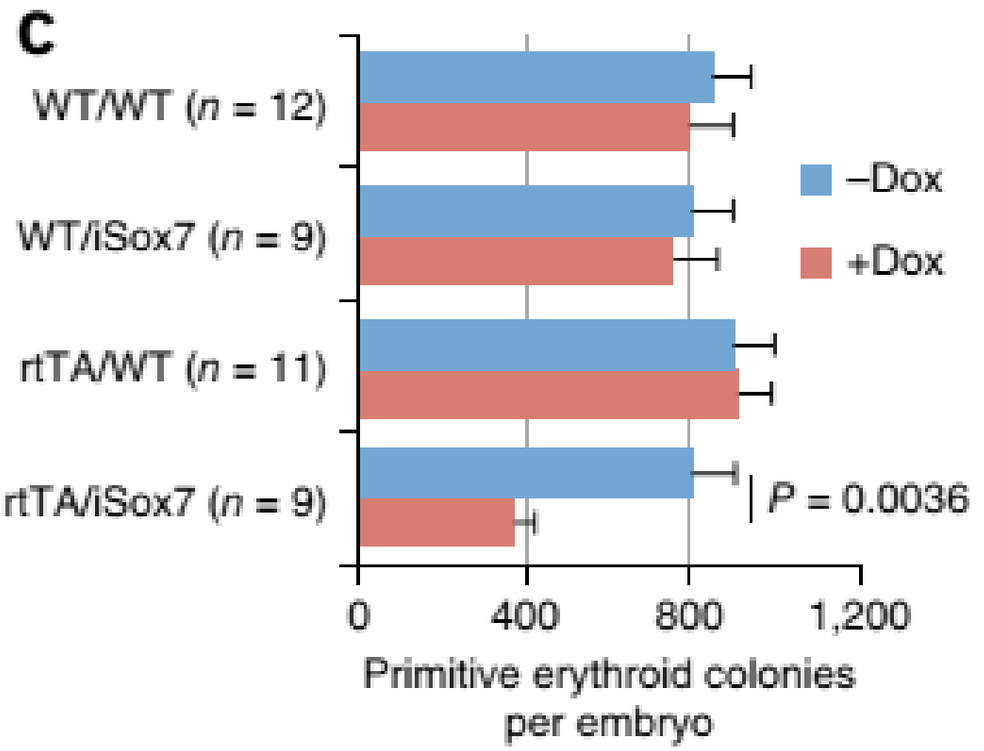
Moignard et al., Nature Biotech. 33, 269 (2015)

(**b**) Colony assays with or without doxycycline from genotyped E8.25 embryos from iSox7^{+}rtTA^{+} mice crossed with wild types.



(**c**) Quantification of primitive erythroid colonies after 4 days.

Embryos carrying both transgenes (bottom) showed a 50% reduction of primitive erythroid colony formation and simultaneous appearance of undifferentiated hemangioblast-like colonies following doxycycline-induced *Sox7* expression compared to controls.
This suggests, in agreement with modeling data and gene expression patterns, that downregulation of *Sox7* is important for the specification of primitive erythroid cells.



In iSox7-mouse, overexpression of Sox7 is stimulated by inducing the Sox7-promoter by addition of the chemical doxycycline (+Dox).

Conclusions

The results indicate, at least for cells destined to become blood and endothelium,
that these cells arise at all stages of the analyzed time course rather than in a synchronized fashion at one precise time point, consistent with the gradual nature of gastrulation.

Using an automated Boolean Network synthesis toolkit we identified a core network of 20 highly connected TFs, which could reach 8 stable states representing blood and endothelium.

We validated model predictions to demonstrate e.g. that Sox7 blocks primitive erythroid development.

Moignard et al.,
Nature Biotech.
33, 269 (2015)