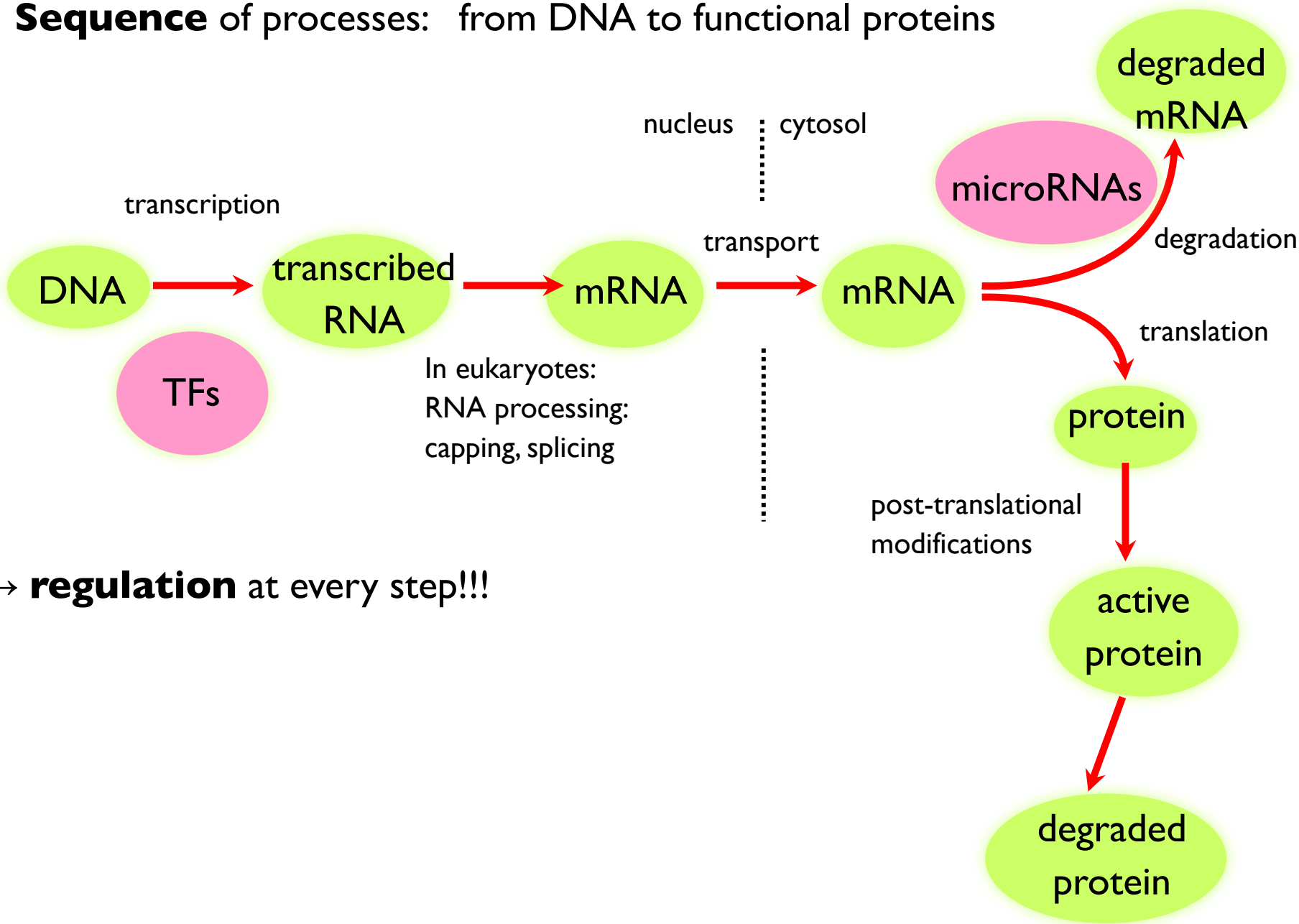
V12 – Gene Regulatory Networks, Boolean Networks

Thu, Nov 28, 2019

Gene Expression

Sequence of processes: from DNA to functional proteins



→ **regulation** at every step!!!

What is a GRN?

Gene regulatory networks (GRN) are model representations of how genes regulate the expression levels of each other.

In **transcriptional regulation**, proteins called **transcription factors (TFs)** regulate the transcription of their **target genes** to produce messenger RNA (mRNA).

In **post-transcriptional regulation, microRNAs** (miRNAs) cause **degradation** and repression of target mRNAs.

These interactions are represented in a GRN by adding edges linking TF or miRNA genes to their target mRNAs.

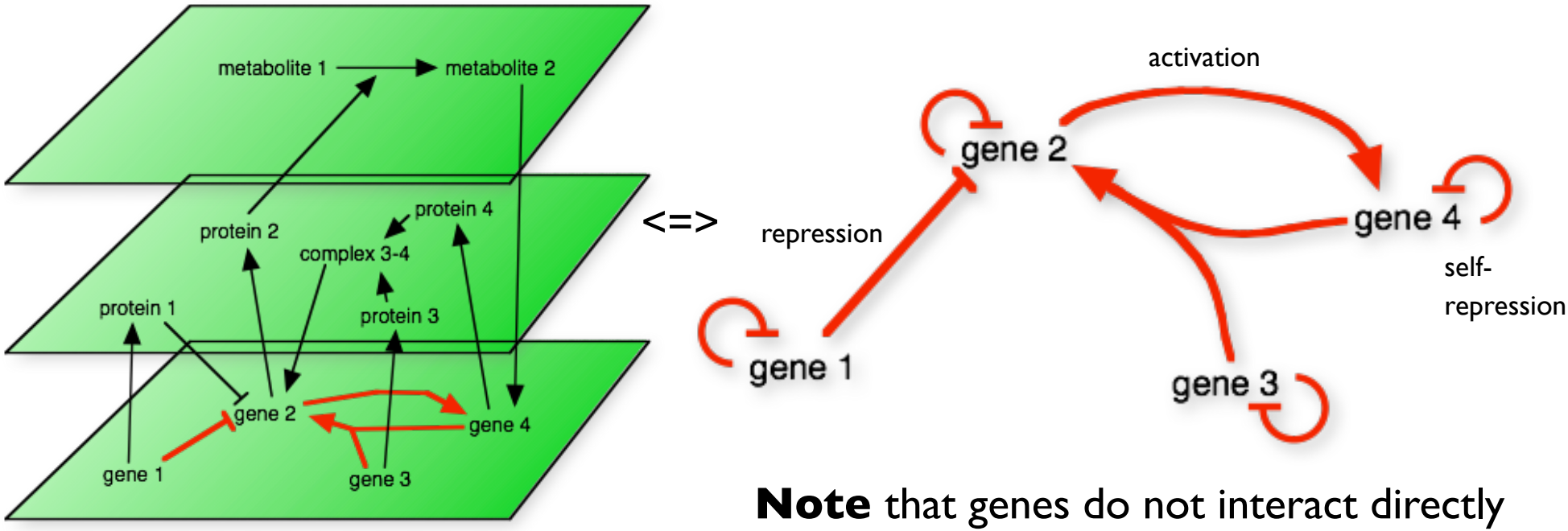
Narang et al. (2015). PLoS Comput Biol 11(9): e1004504

Layers upon Layers

Biological regulation
via proteins and metabolites

<=>

Projected regulatory network



Note that genes do not interact directly

Gene regulation networks have "cause and action"
→ **directed** networks

A gene can enhance or suppress the expression of another gene
→ **two types** of arrows

Global Regulators in *E. coli*

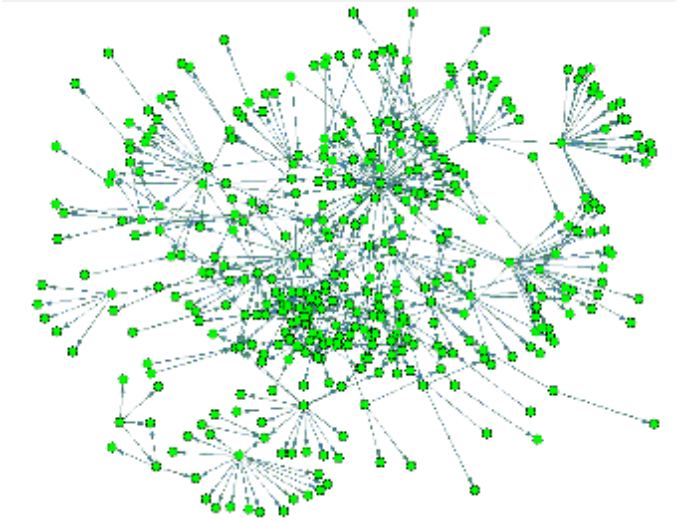
Table 1: Global regulators and their regulated operons and functions in the regulatory network of *E. coli*.

Global regulator	directly regulated Operons	Total regulated operons	Modules regulated	Function
IHF	21	39	15	integration host factor
CspA	2	24	5	Cold shock protein
CRP	72	112	21	cAMP receptor protein
FNR	22	38	16	anaerobic regulator, regulatory gene for nitrite and nitrate reductases, fumarate reductase
HNS	7	22	5	DNA-binding global regulator; involved in chromosome organization; preferentially binds bent DNA
OmpR	6	20	3	Response regulator for osmoregulation; regulates production of membrane proteins
RpoN	12	17	4	RNA polymerase sigma 54 subunit
RpoS	14	24	8	stationary phase sigma factor
ArcA	20	21	6	Response regulator protein represses aerobic genes under anaerobic growth conditions and activates some anaerobic genes
NarL	13	15	5	Two-component regulator protein for nitrate/nitrite response

Ma et al., *BMC Bioinformatics* **5** (2004) 199

Simple organisms have hierarchical GRNs

Largest weakly connected component (WCC)
(ignore directions of regulation): 325 operons
(3/4 of the complete network)

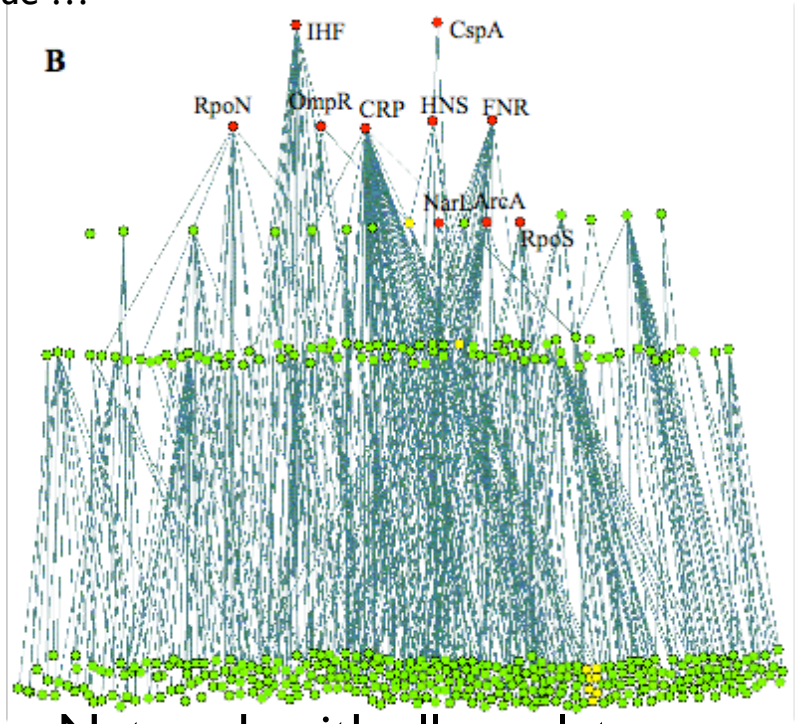


Network from standard layout algorithm

→

Lowest level: operons that code for TFs with only auto-regulation, or no TFs

Next layer: delete nodes of lower layer, identify TFs that do not regulate other operons in this layer (only lower layers)
Continue …



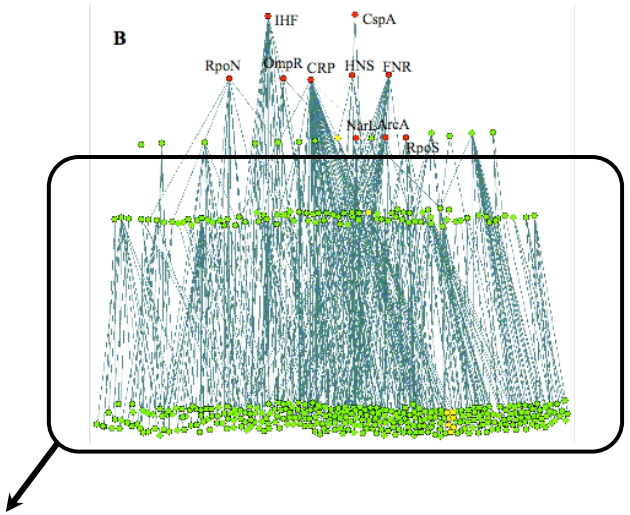
Network with all regulatory edges pointing downwards

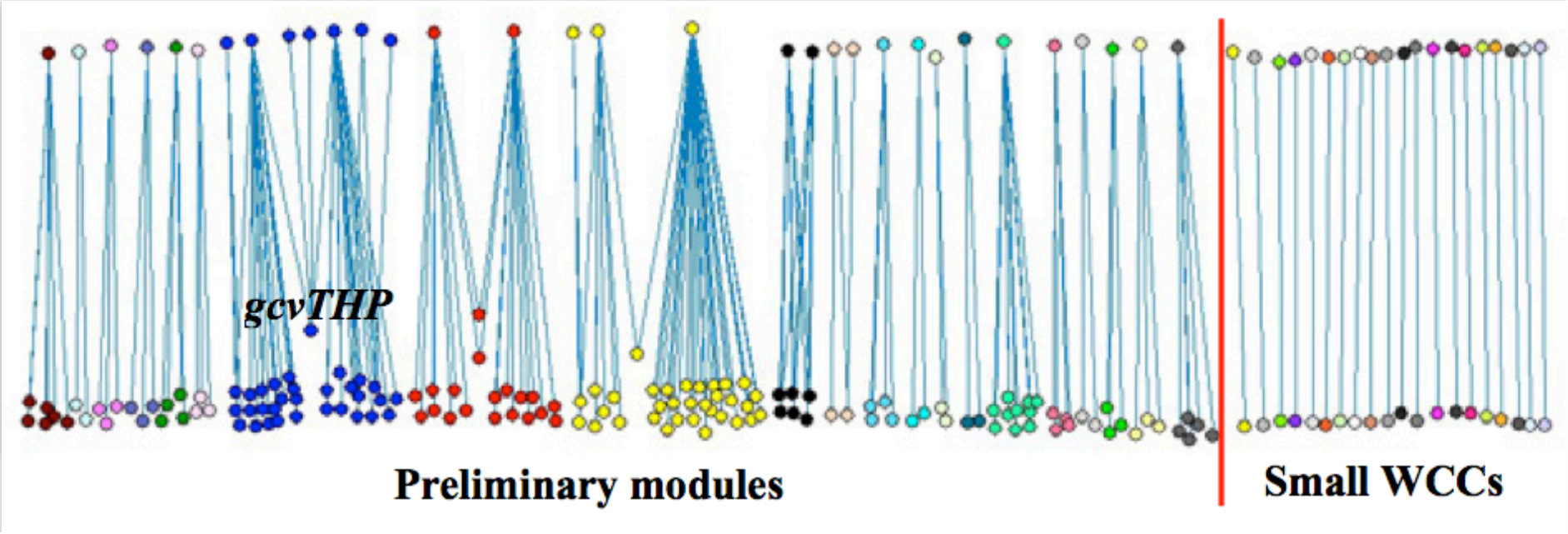
→ a few global regulators (•) control all the details

Ma et al., *BMC Bioinformatics* **5** (2004) 199

E.coli GRN modules

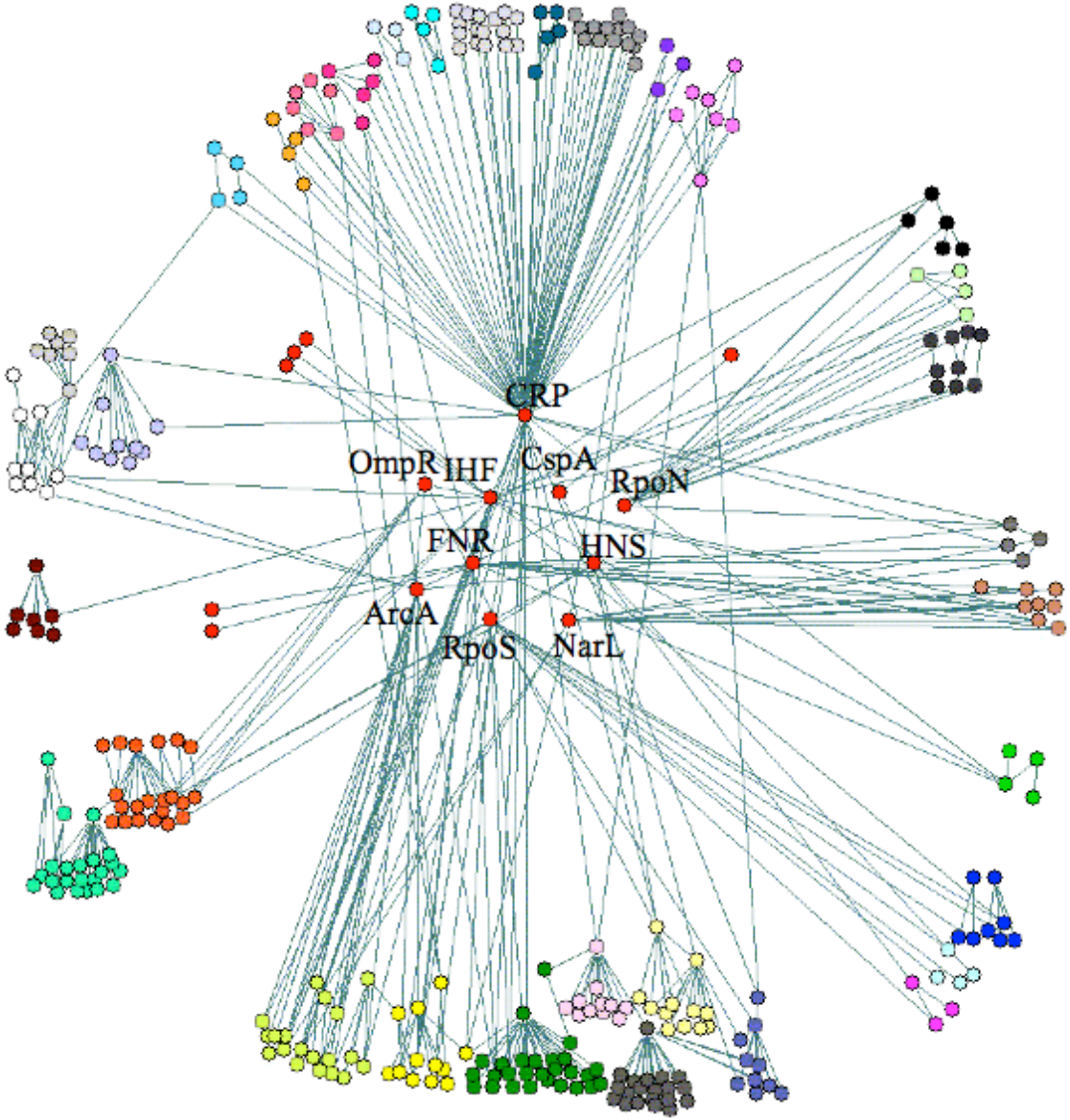
Remove top 3 layers and determine WCCs
→ just a few modules





Ma et al., *BMC Bioinformatics* **5** (2004) 199

Putting it back together



The 10 global regulators are at the core of the network,

some hierarchies exist between the modules

Ma et al., *BMC Bioinformatics* **5** (2004) 199

Modules have specific functions

Table 2: Functional investigation of modules identified.

index	Operons included	Biological function description
1	*aceBAK, acs, adhE, fruBKA, fruR, icdA, iclMR, mlc, ppsA, ptsG, ptsHI_crr, pykF*	Hexose PTS transport system, PEP generation, Acetate usage, glyoxylate shunt
2	*acnA, fpr, fumC, marRAB, nfo, sodA, soxR, soxS, zwf*	Oxidative stress response
3	*ada_alkB, aidB, alkA, ahpCF, dps, gorA, katG, oxyR*	Oxidative stress response, Alkylation
4	*alaWX, aldB, argU, argW, argX_hisR_leuT_proM, aspV, dnaA, leuQPV, leuX, lysT_valT_lysW, metT_leuW_glnUW_metU_glnVX, metY_yhbC_nusA_infB, nrdAB, pdhR_aceEF_lpdA, pheU, pheV, proK, proL, proP, sdhCDAB_b0725_sucABCD, serT, serX, thrU_tyrU_glyT_thrT, thrW, tyrTV, valUXY_lysV, yhdG_fis*	rRNA, tRNA genes, DNA synthesis system, pyruvate dehydrogenase and ketoglutarate dehydrogenase system
5	*araBAD, araC, araE, araFGH, araJ*	Arabinose uptake and usage
6	*argCBH, argD, argE, argF, argI, argR, carAB*	Arginine usage, urea cycle
7	*caiF, caiTABCDE, fixABCX*	Carnitine usage
8	*clpP, dnaKJ, grpE, hflB, htpG, htpY, ibpAB, lon, mopA, mopB, rpoH*	Heat shock response
9	*codBA, cvpA_purF_ubiX, glnB, glyA, guaBA, metA, metH, metR, prsA, purC, purEK, purHD, purL, purMN, purR, pyrC, pyrD, speA, ycfC_purB, metC, metF, metJ*	Purine synthesis, purine and pyrimidine salvage pathway, methionine synthesis
10	*cpxAR, cpxP, dsbA, ecfI, htrA, motABcheAW, ppiA, skp_lpxDA_fabZ, tsr, xprB_dsbC_recJ*	Stress response, Conjugative plasmid expression, cell motility and Chemotaxis
11	*dctA, dcuB_fumB, frdABCD, yjdHG*	C4 dicarboxylate uptake
12	*edd_eda, gntKU, gntR, gntT*	Gluconate usage, ED pathway
13	*csgBA, csgDEFG, envY_ompT, evgA, gcvA, gcvR, gcvTHP, gltBDF, ilvIH, kbl_tdh, livJ, livKHMGF, lrp, lysU, ompC, ompF, oppABCDF, osmC, sdaA, serA, stpA*	Amino acid uptake and usage
14	*fdhF, fhlA, hycABCDEFGH, hypABCDE*	Formate hydrogenlyase system
15	*flgAMN, flgBCDEFGHIJ, flgKL, flgMN, flhBAE, flhDC, fliAZY, fliC, fliDST, fliE, fliFGHIJK, fliLMNOPQR, tarTapcheRBYZ*	Flagella motility system
16	*ftsQAZ, rcsAB, wza_wzb_b2060_wcaA_wcaB*	Capsule synthesis, cell division
17	*gdhA, glnALG, glnHPQ, nac, putAP*	Glutamine and proline utilization
18	*glmUS, manXYZ, nagBACD, nagE*	Glucosamine, mannose utilization
19	*glpACB, glpD, glpFK, glpR, glpTQ*	Glycerol phosphate utilization
20	*lysA, lysR, tdcABCDEFG, tdcR*	Serine, threonine usage
21	*malEFG, malK, lamB, malM, malPQ, malS, malT, malZ*	Maltose utilization

Ma et al., *BMC Bioinformatics* **5** (2004) 199

Frequency of co-regulation

Half of all target genes are regulated by multiple TFs.

In most cases, a „gobal“ regulator (with > 10 interactions) works together with a more specific local regulator.

Martinez-Antonio, Collado-Vides, Curr Opin Microbiol 6, 482 (2003)

Table 1

Summary of transcriptional interactions of major TFs, in the transcriptional regulatory network of *E. coli*.

Transcription factor	Genes regulated*	Co-regulators†	TFs regulated‡	Sigma factors§	Functional classes of genes regulated#	Family (members)¶
CRP	197	47	22	$\sigma^{70,38,32,24}$	48	CRP (2)
IHF	101	28	9	$\sigma^{70,54,38}$	26	HI-HNS (2)
FNR	111	20	5	$\sigma^{70,54,38}$	22	CRP (2)
FIS	76	15	4	$\sigma^{70,38,32}$	20	EBP (14)
ArcA	63	18	2	$\sigma^{70,38}$	17	OmpR (14)
Lrp	53	14	3	$\sigma^{70,38}$	15	AsnC (3)
Hns	26	14	5	$\sigma^{70,38,32}$	17	Histone-like (1)
NarL¥	65	10	1	$\sigma^{70,38,}$	14	LuxR/UhpA (17)
OmpR	10	9	3	$\sigma^{70,38}$	5	OmpR (14)
Fur¥	26	8	2	$\sigma^{70,19}$	9	Fur (2)
PhoB	26	1	3	σ^{70}	9	OmpR (14)
CpxR	9	2	1	$\sigma^{70,38,24}$	5	OmpR (14)
SoxRS	9	10	3	$\sigma^{70,38}$	10	AraC/XylS (24)
Mlc¥	5	3	1	$\sigma^{0,32}$	3	NagC/XylR (7)
CspA¥	2	2	1	σ^{70}	2	Cold (9)
Rob**	7	8	2	$\sigma^{70,38}$	6	AraC/XylS (27)
PurR**	28	7	1	σ^{70}	10	GalR/LacI (13)

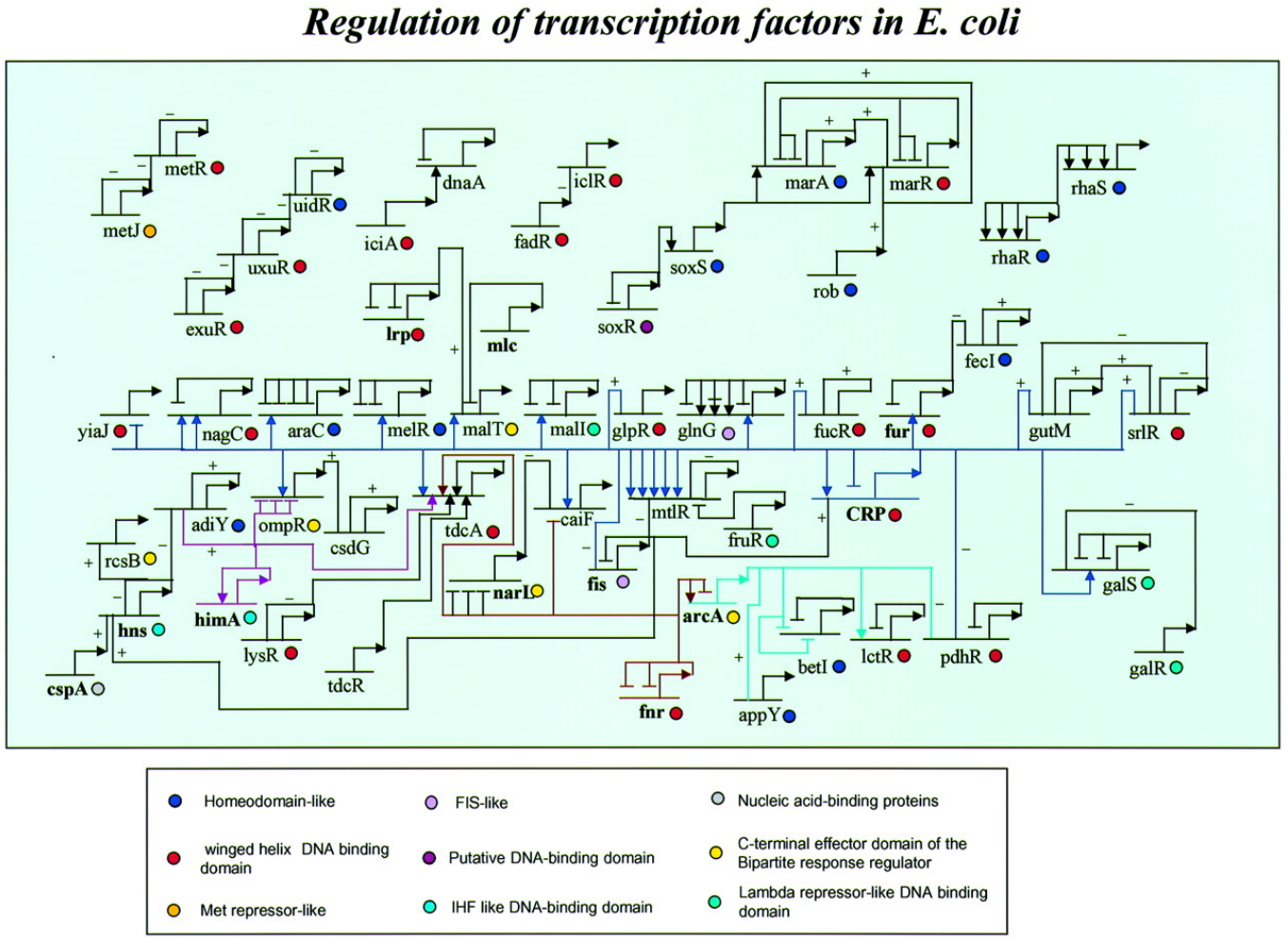
*Total number of genes regulated directly. †Number of different TFs with which at least a gene or TU is jointly co-regulated. ‡Number of regulated genes that codify for TFs. §List of σ factors of the regulated promoters. #Number of functional classes of the gene products regulated [44]. ¶TF family and in parenthesis the number of members of the family. In addition to the seven global TFs considered here there are TFs suggested by ¥Babu and Teichmann, 2003, [42••] and **Shen-Orr *et al.*, 2002, [50••].

TF regulatory network in *E.coli*

When more than one TF regulates a gene, the order of their binding sites is as given in the figure.

Arrowheads and **horizontal bars** indicate positive / negative regulation when the position of the binding site is known.

In cases where only the nature of regulation is known, without binding site information, + and – are used to indicate positive and negative regulation.

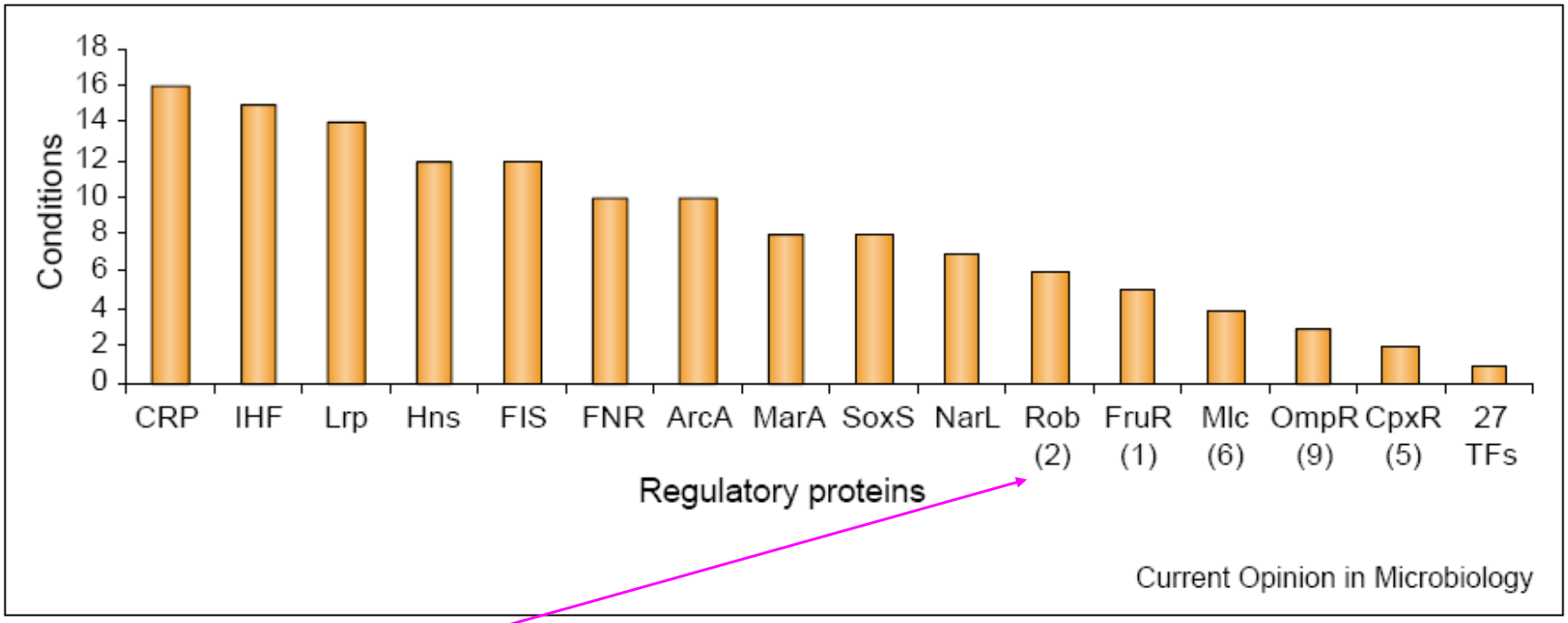


The names of **global regulators** are in **bold**.

Babu, Teichmann, Nucl. Acid Res. 31, 1234 (2003)

Response to changes in environmental conditions

TFs also sense changes in environmental conditions or other changes that encode internal signals.



Global environment growth conditions in which TFs are regulating.
in brackets indicates how many additional TFs participate in the same number of conditions.

Martinez-Antonio, Collado-Vides, Curr Opin Microbiol 6, 482 (2003)

Story: Quorum sensing of *Vibrio fischeri*

V. fischeri has a microbial **symbiotic relationship** with the squid *Euprymna scolopes*.
The bacterium exists in **small amounts** in the ocean (10^2 cells/ml) and in **large amount** in the light organs of the **squid** (10^{10} cells/ml).

At low concentrations, *V. fischeri* does not produce luminescence.

At **high cell density** these bacteria emit a **blue-green light**.

The light organ of the squid provides to the bacteria all the **nutrients** that they need to survive.

The squid benefits from the bacteria's quorum sensing and **bioluminescence** abilities.

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

Quorum sensing of *Vibrio fischeri*

The cell density-dependent control of gene expression is activated by a transcriptional activator protein that is coupled to a signal molecule (**autoinducer**).

The autoinducer is released by the bacteria into its surrounding **environment** and taken up from there.

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.

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

Vibrio fischeri helps with Camouflage

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

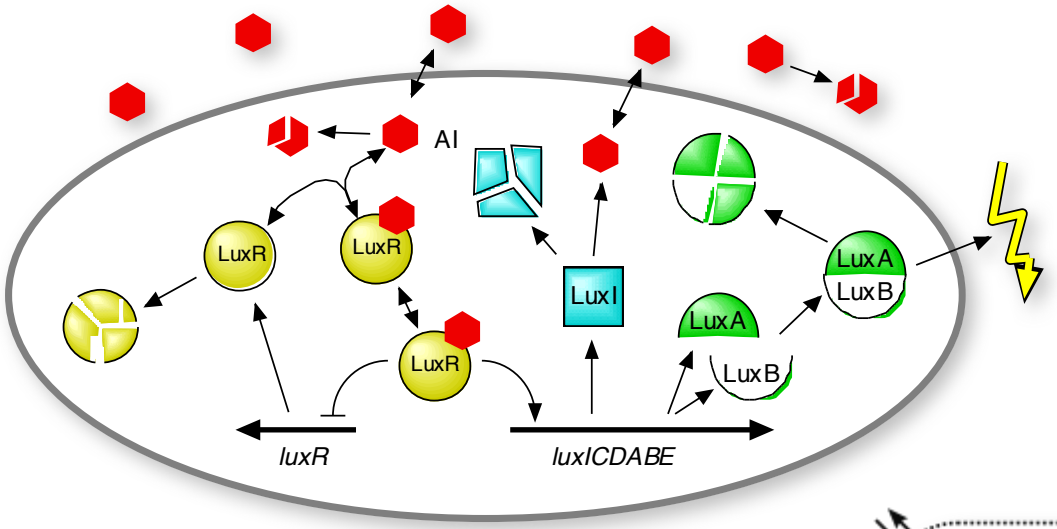
In the **moonlight**, the swimming squid would normally cast a **shadow** beneath itself making it a perfect target for squid-eating organisms.

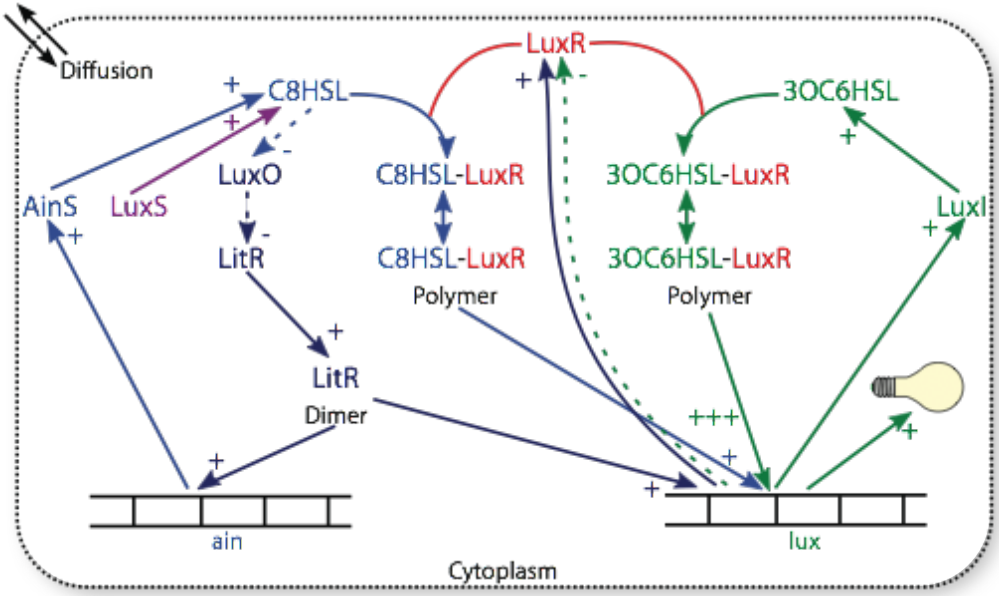
However, the bacterial glow will counter the shadowing effect the moon makes and mask the squid from its predators.

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

Dependencies between variables can be 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**

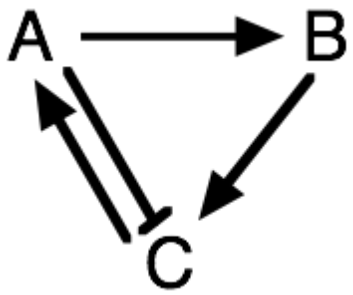
→ all states leading to an attractor = **basin of attraction**

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

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

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

Test the Other Starting Conditions

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

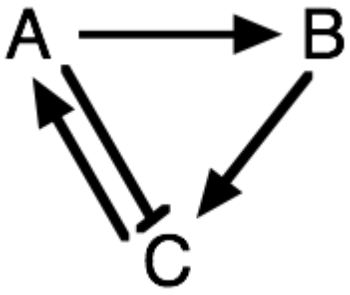
#	A	B	C
0	0	1	1
1	1	0	1

Same attractor as before:

100 → 010 → 001 → 100

is also reached from:

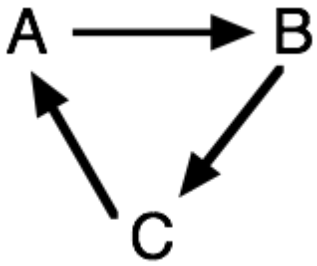
110, 111, 101, 011



#	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

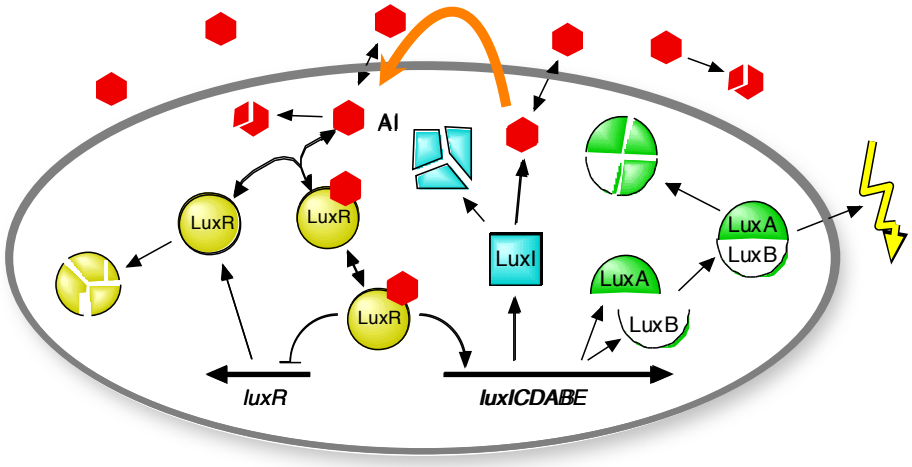
#	A	B	C
0	1	1	1
1	1	1	1

#	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.

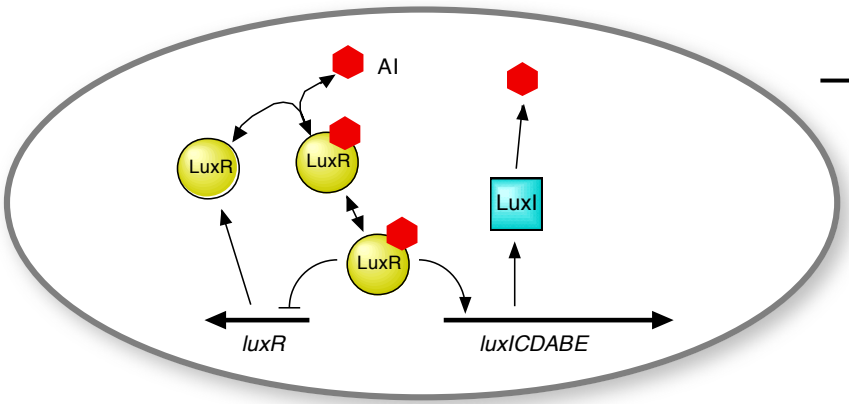
LuxI	LuxR:AI:Genome
0	0
1	1

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

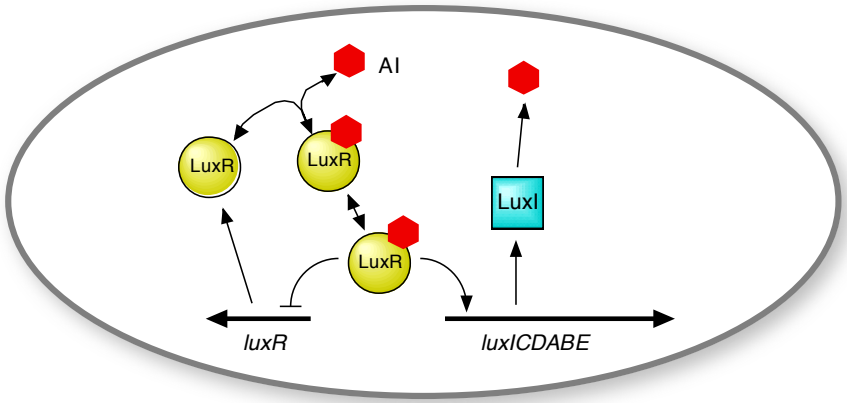
LuxR:AI	LuxR	AI	LuxR:AI:Genome
0	0	0	0
0	I	0	0
0	0	I	0
I	I	I	0
0	0	0	I
0	I	0	I
0	0	I	I
I	I	I	I

→

LuxR:AI	LuxR	AI	LuxR:AI:Genome
0	x	x	x
I	I	I	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	I	0	0
I	0	I	0
0	I	I	0
I	0	0	I
I	I	0	I
I	0	I	I
I	I	I	I

→

AI	LuxR	AI	LuxI
I	x	x	I
0	x	0	0
I	0	I	0
0	I	I	0

Scanning for Attractors

States of *V. fischeri* QS system are 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 up 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

States: named by reading occupancies as **binary numbers in reversed order**.

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:
17 returns to 5

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
	.	X	X	.	.
	.	X	X	X	.
	.	.	X	X	X
	.	.	X	.	X

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

#	LR	RA	AI	RAG	LI
	X	X	X	.	.
	X	X	.	X	.
	X	.	.	X	X
	X	.	X	.	X

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

#	LR	RA	AI	RAG	LI
	X	.	X	X	.
	.	X	.		X

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

There exist three **regimes**:

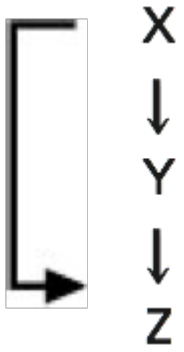
dark: LuxI = 0 — free LuxR, no AI

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

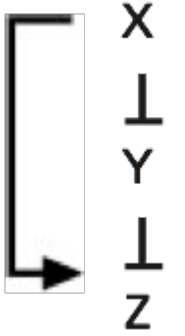
External signal determines state of X
→ response Z for short and long signals X



condition tables:

Y	X
0	0
I	I

Z	X	Y
0	0	0
0	0	I
0	I	0
I	I	I



Y	X
I	0
0	I

Z	X	Y
0	0	0
0	0	I
I	I	0
0	I	I

Signal propagation

Left column: external signal

Response to signal X(t)

X	Y	Z
0	0	0
I	0	0
0	I	0
0	0	0
I	0	0
I	I	0
I	I	**I**
0	I	**I**
0	0	**0**
0	0	0

Short Signal

Long signal

X	Y	Z
0	I	0
I	I	0
0	0	0
0	I	0
I	I	0
I	0	0
I	0	**I**
0	0	**I**
0	I	**I**
0	I	0

Can Boolean Networks be predictive?

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 → dynamic balances lost, reduced to oscillations

• conditional yes–no causality → no continuous processes

• discretized propagation steps → timing of concurrent paths?

"You get what you pay for"

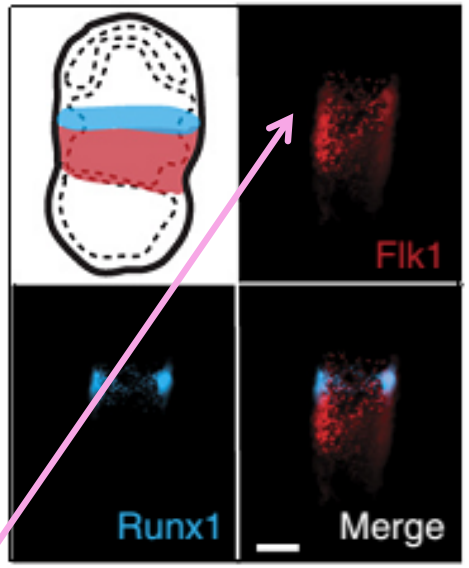
Understand Blood development (hemato-poeisis) with the help of Boolean Networks

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

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.)

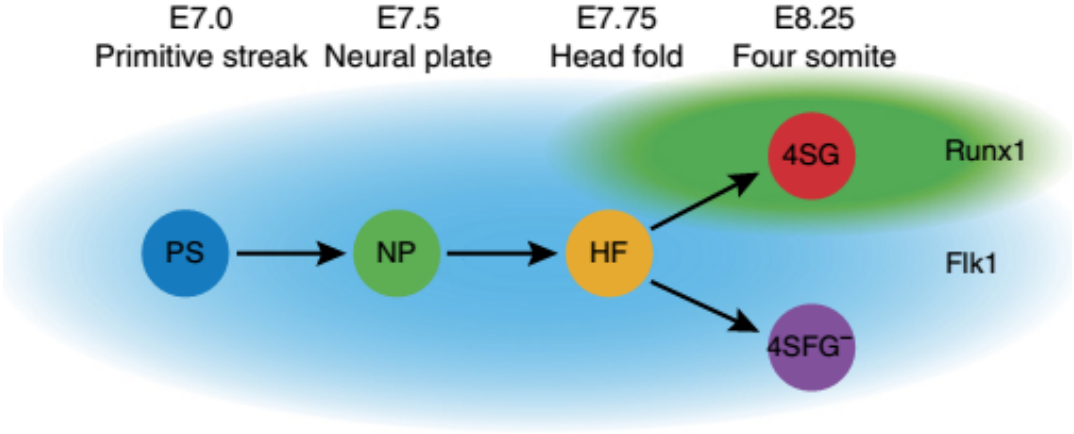
These cells initially have the potential to form either blood, endothelium and smooth muscle cells.



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

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.

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

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

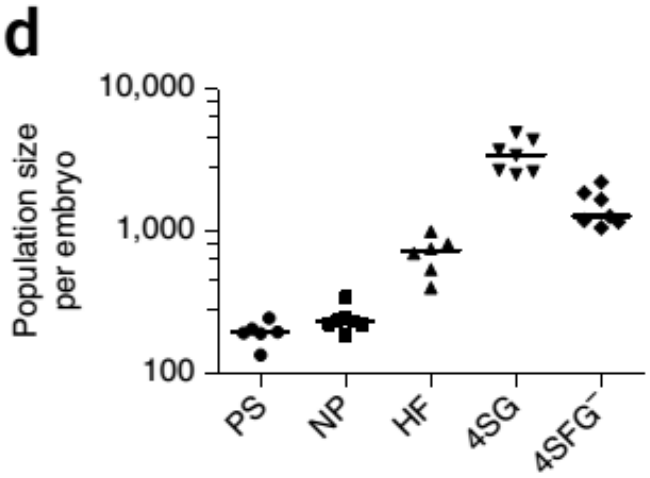
Moignard et al., Nature Biotech. 33, 269 (2015)

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 different stages present in each embryo were estimated from fluorescence-activated cell sorting (FACS) data.



Number of cells grows as embryonic development progresses.

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

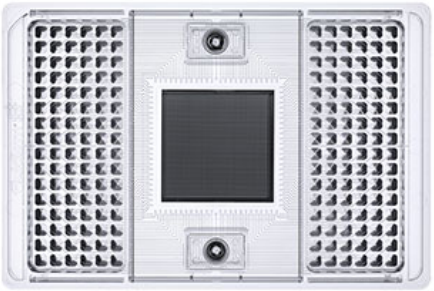
Assay gene expression in single cells

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:

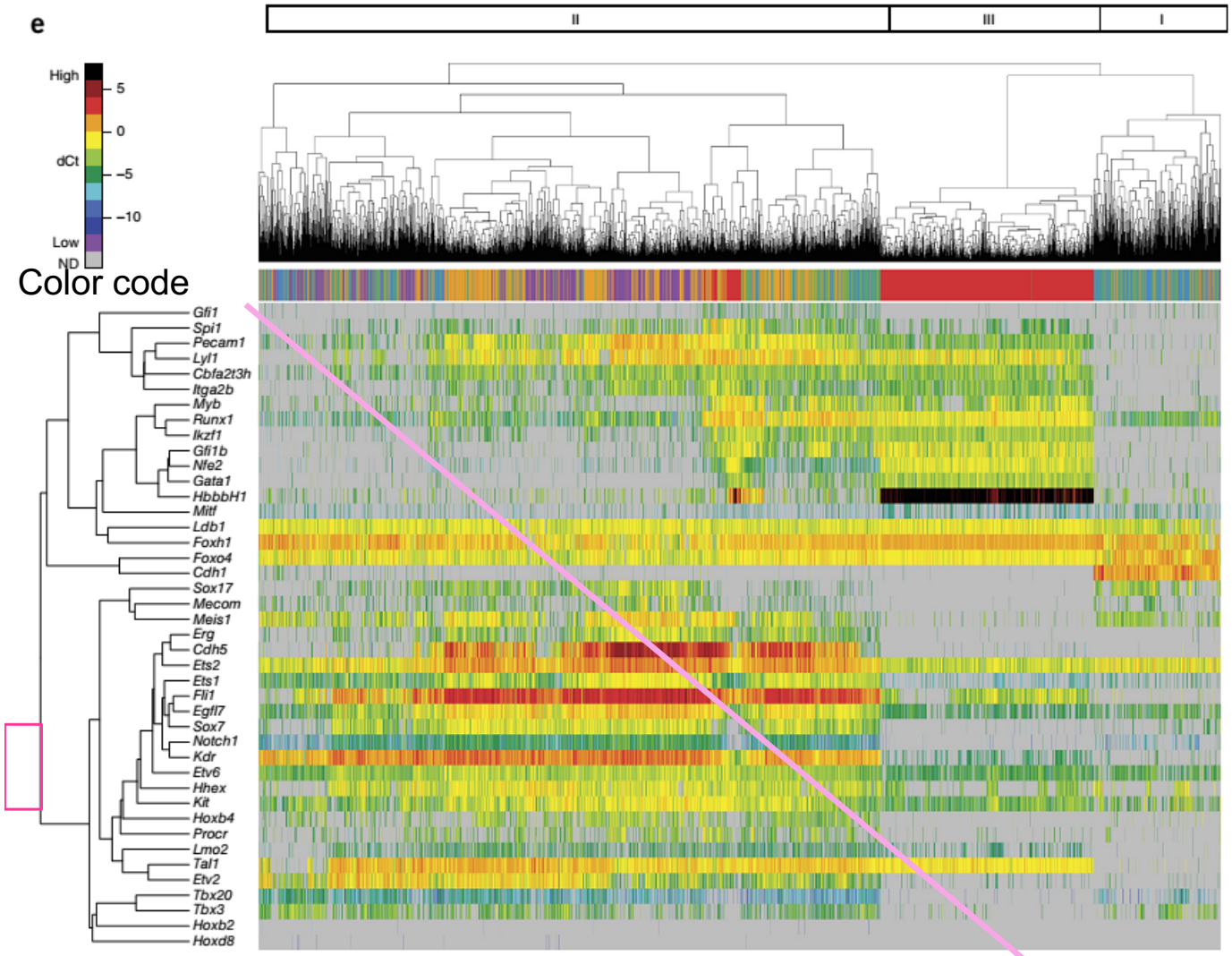
- 33 **transcription factors** known to be involved in endothelial and hematopoietic development
- 9 **marker genes** (needed for FACS-sorting)
- 4 **house-keeping genes** (needed for quality checks and normalization)



www.fluidigm.com

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

Hierarchical clustering of gene expression data



3 main clusters:

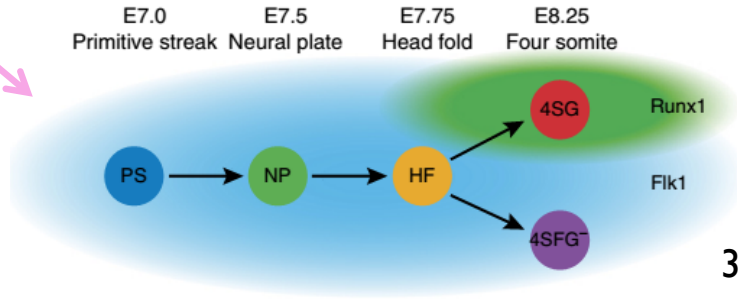
Cluster I (right side) contains mostly PS and NP cells (**green**/**blue**)

Cluster III contains exclusively 4SG cells (**red**)

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

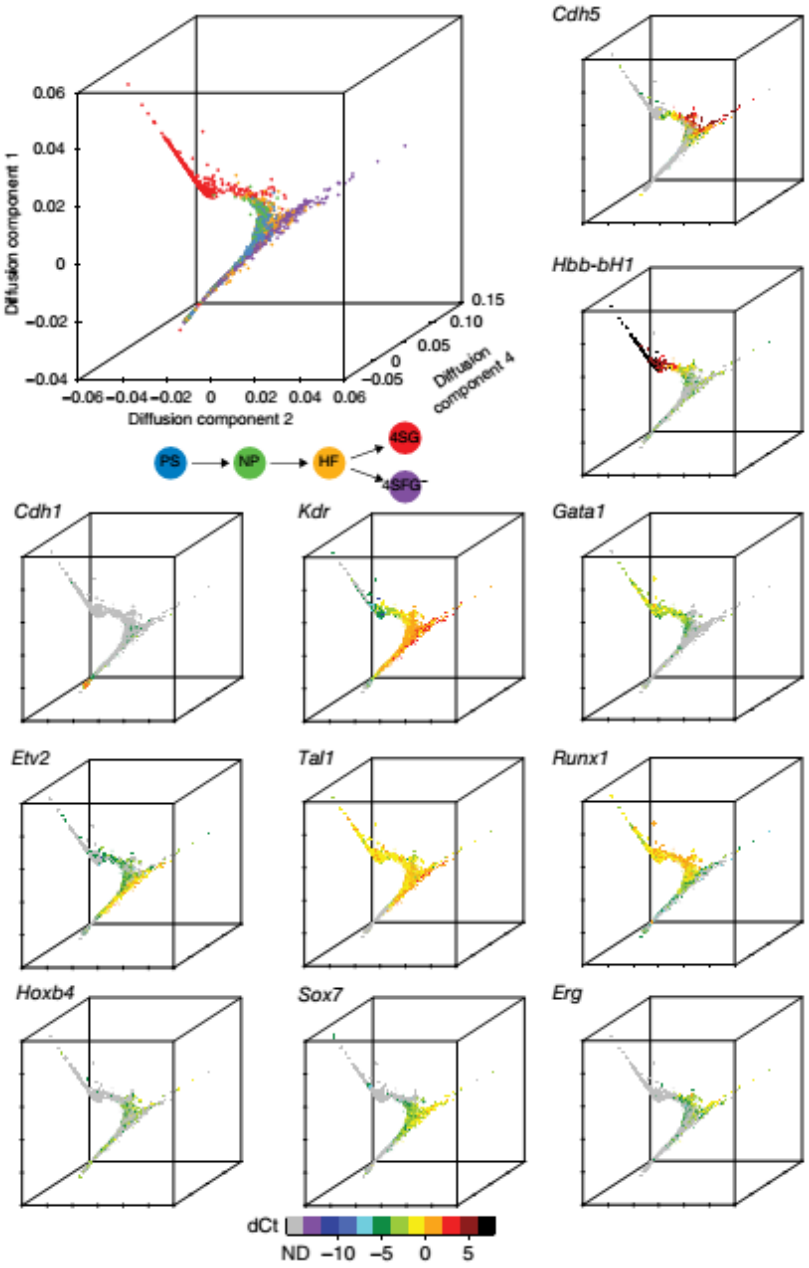
→ Cell differentiation progresses asynchronously

← Single cells →



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)

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

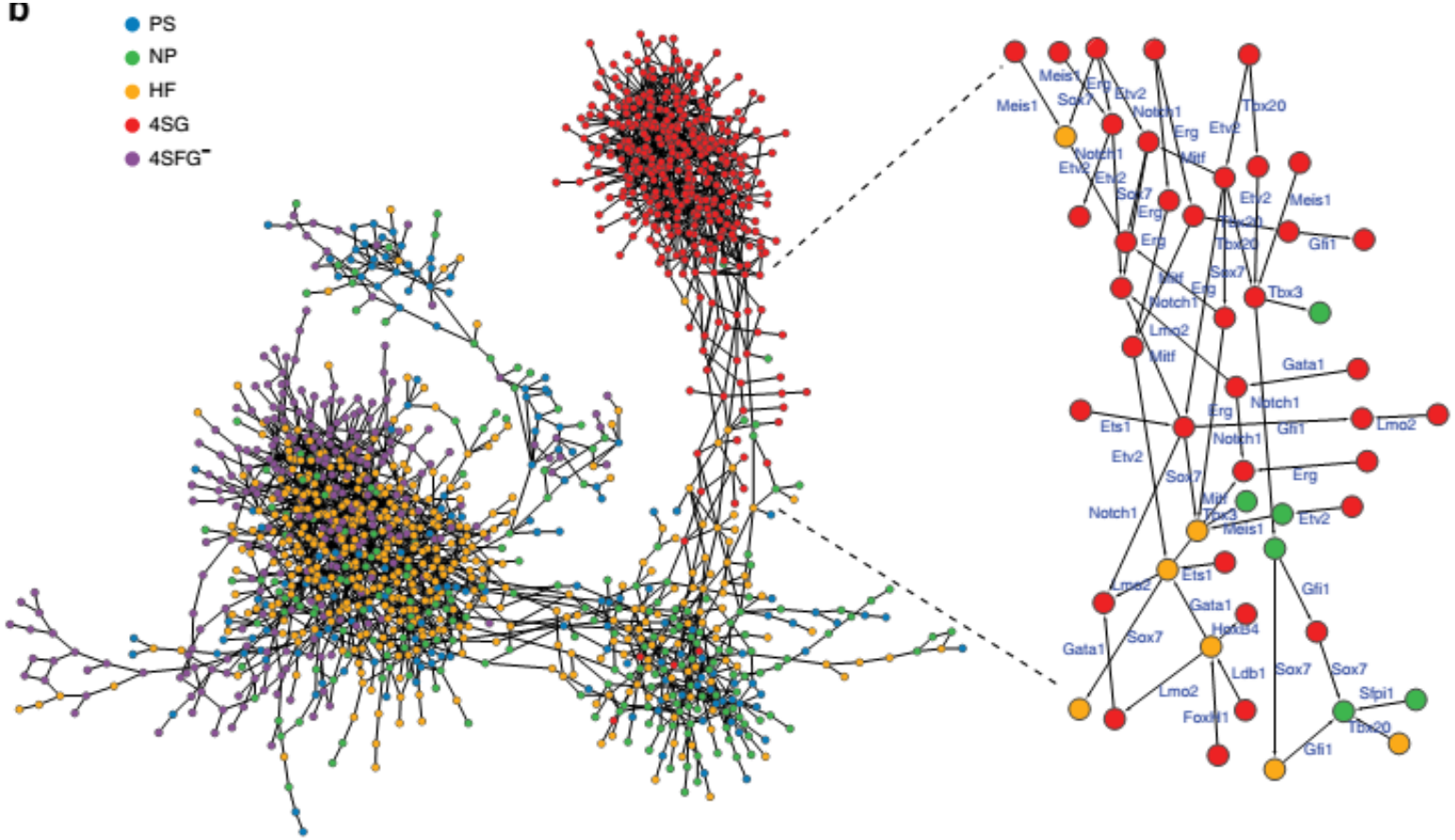
Determine suitable **expression thresholds** for each gene to categorize its expression levels into **binary** on / off states.

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

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

State graph of largest connected comp.

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



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

Add **edges** to connect all those pairs of states that differ in the on/off levels of a single gene (and are identical otherwise), see right side with labeled edges.

Idea behind this: these transitions can be best interpreted.

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

The rule 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 **is maximized** in which no transitions induced by the update functions are **missing**.

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

The update function u_i is restricted 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

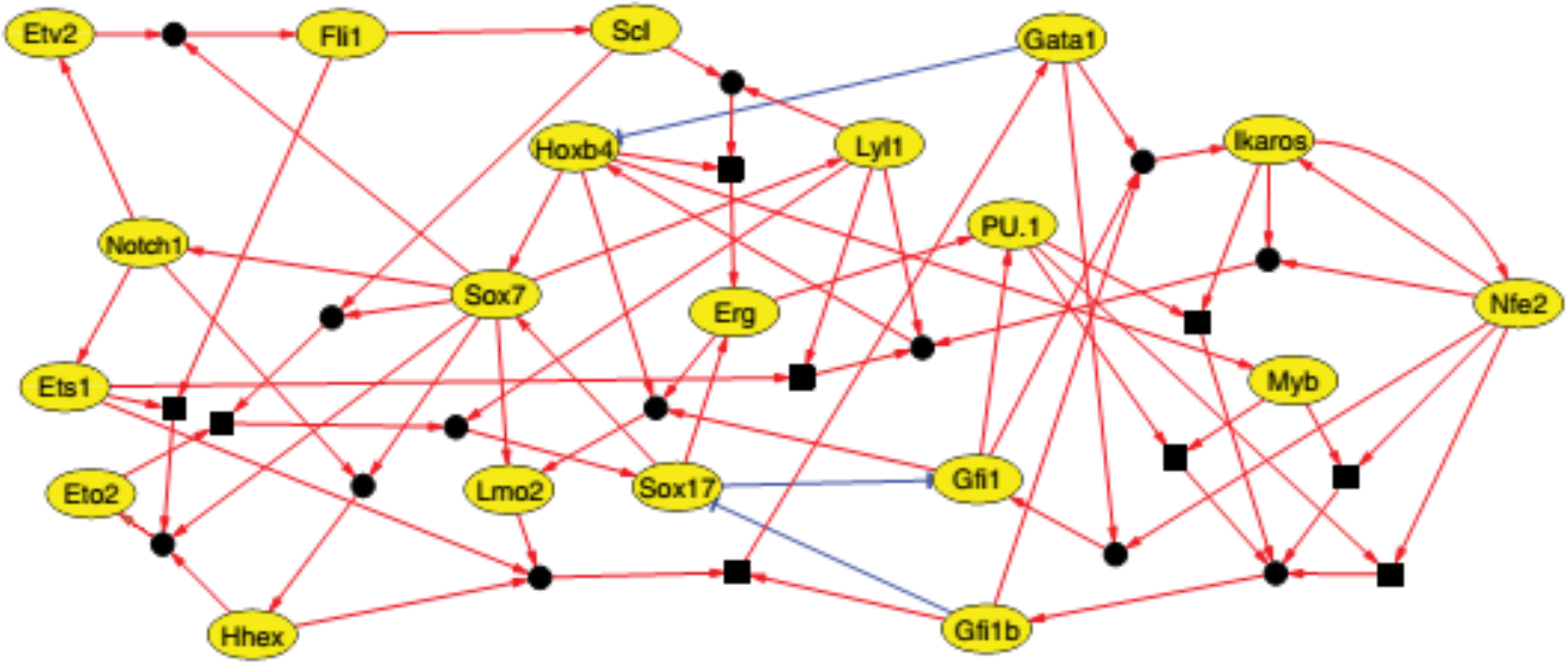
Additional validity check of the postulated rules:

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

Simulate overexpression and knockout experiments for each TF.

Assess ability of the network to reach wildtype or new stable states.

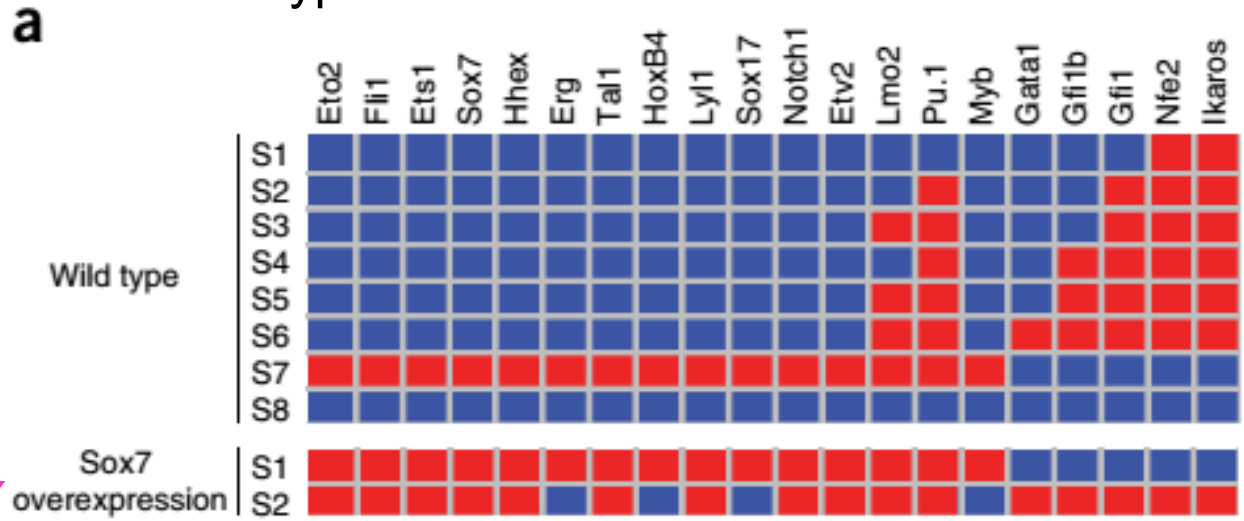
Red : gene expressed;

blue : gene not expressed.

S2-S6: blood-like

S7: endothelial-like

S8 : no activity



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

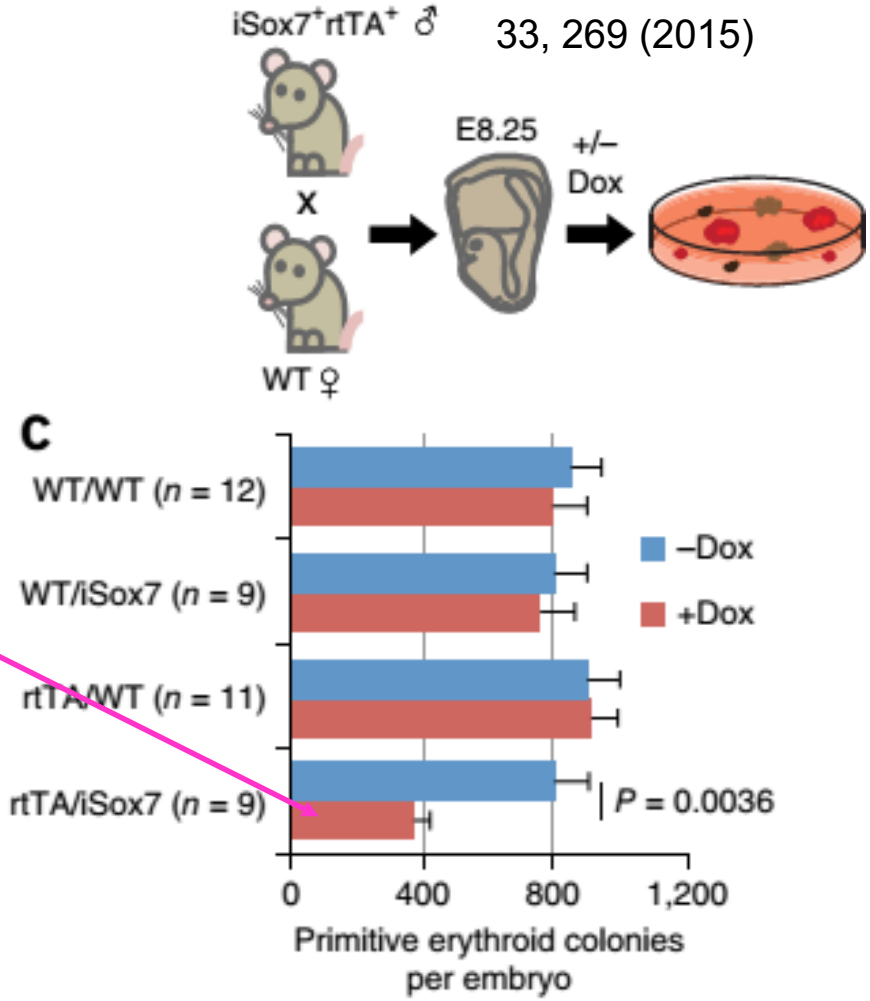
Moignard et al., Nature Biotech. 33, 269 (2015)

(**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 (rtTA/iSox7) showed a **50% reduction of primitive erythroid colony formation** 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

Cells destined to become blood and endothelium arise at all stages of the analyzed time course rather than in a synchronized fashion at one precise time point. This is consistent with the gradual nature of gastrulation.

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

The model predictions could be validated by demonstrating e.g. that Sox7 blocks primitive erythroid development.

→ Boolean Networks can be predictive and may guide experiments.

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