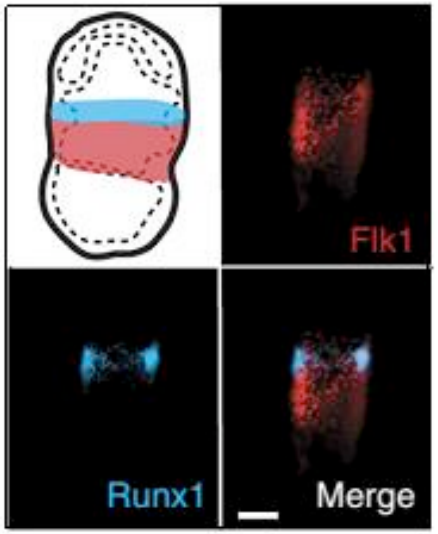
V8: Hematopoeisis

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 initiates at **gastrulation** from multipotent Flk1^{+} mesodermal cells, which initially have the potential to form blood, endothelium and smooth muscle cells.

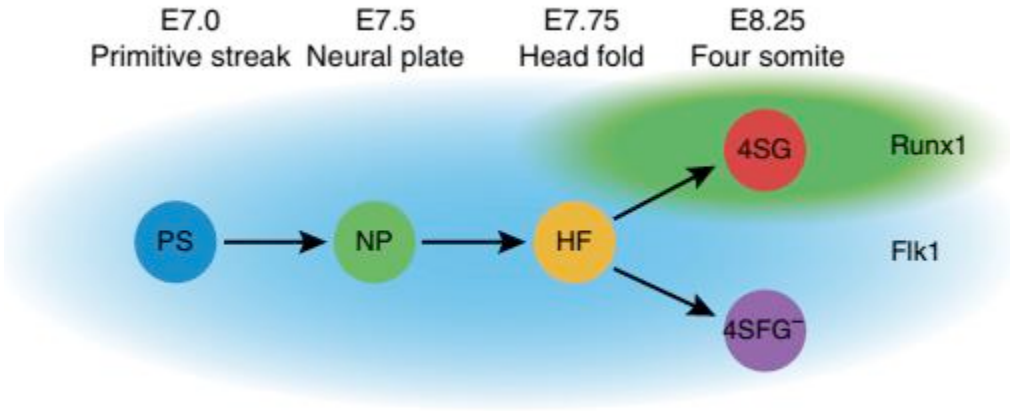
Blood development represents one of the **earliest stages** of **organogenesis**, as the production of primitive erythrocytes is required to support the growing embryo.



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.

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

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

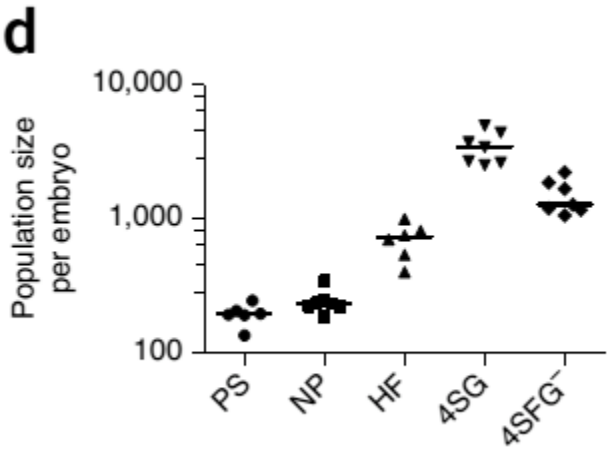
Moignard et al.,
Nature Biotech.
33, 269 (2015)

Material

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.



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

What experiments should be performed

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.

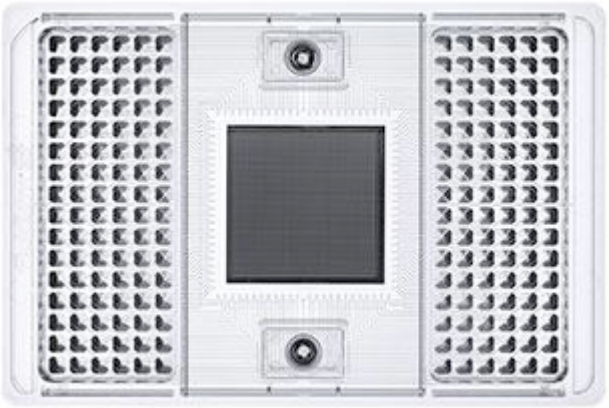
Genes essayed

- 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)

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

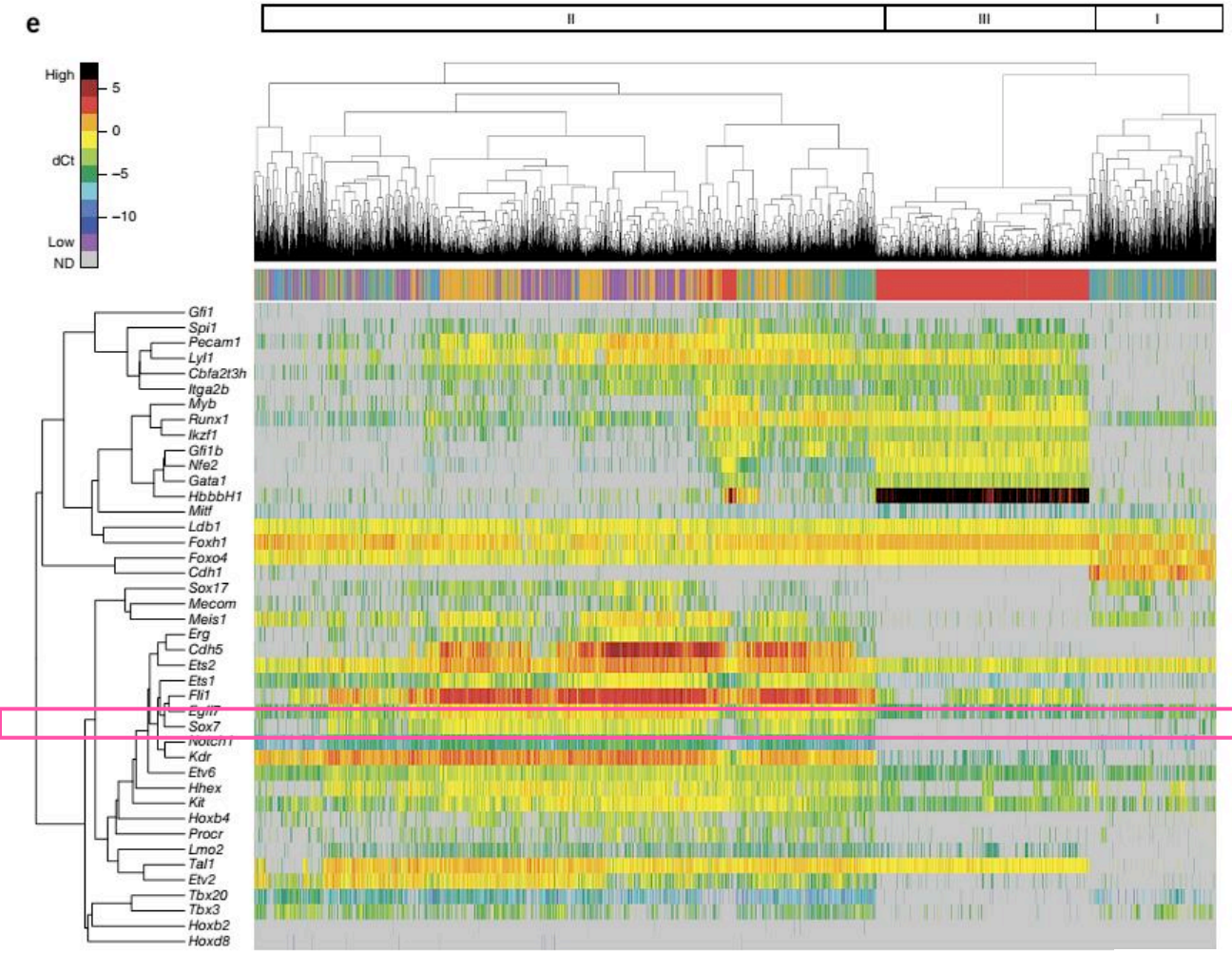
Fluidigm biomark: collect gene expression in single cells

“Fluidigm’s revolutionary integrated fluidic circuits (IFCs) empower life science research by automating PCR reactions in nanoliter volumes.”



www.fluidigm.com

Hierarchical clustering of gene expression data



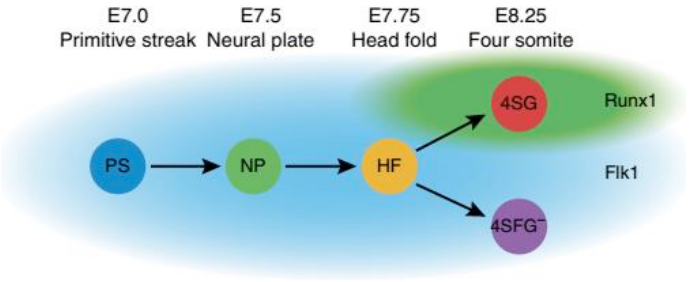
3 main clusters:

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

Cluster III contains exclusively 4SG cells

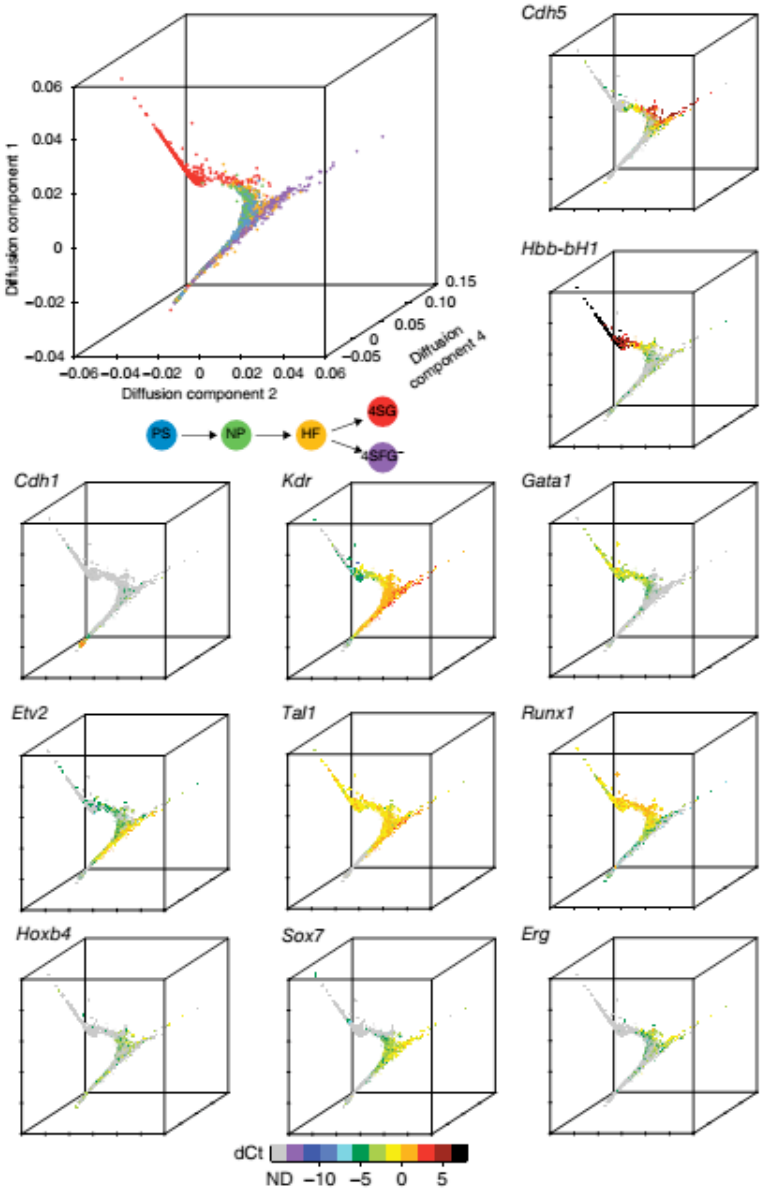
Cluster II is mixed (NF, 4SFG$^-$, …)

→ 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 $\sum_{i=1} P(i,j) = 1$

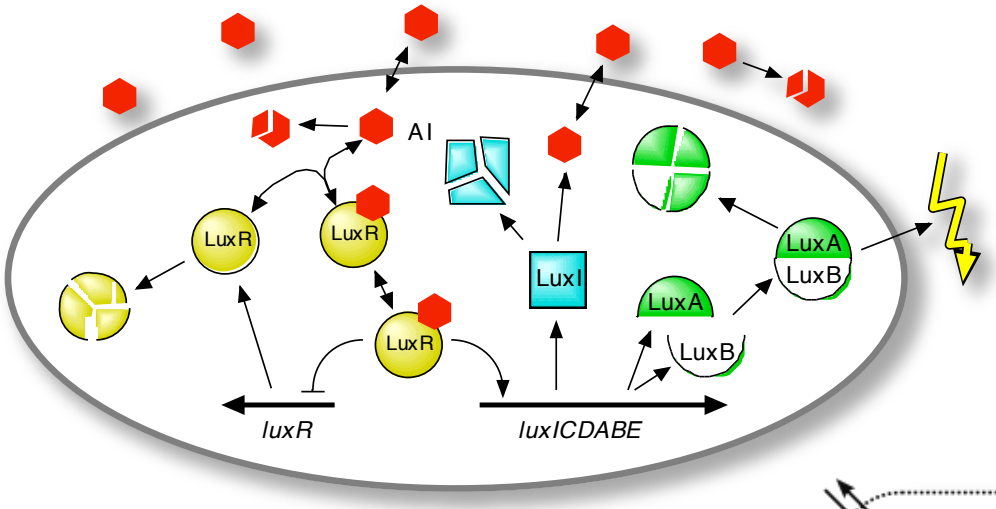
The cells are organized in 2D or 3D such that the Euclidean distance between the cells corresponds to the diffusion metric *P*(*i*,*j*) .

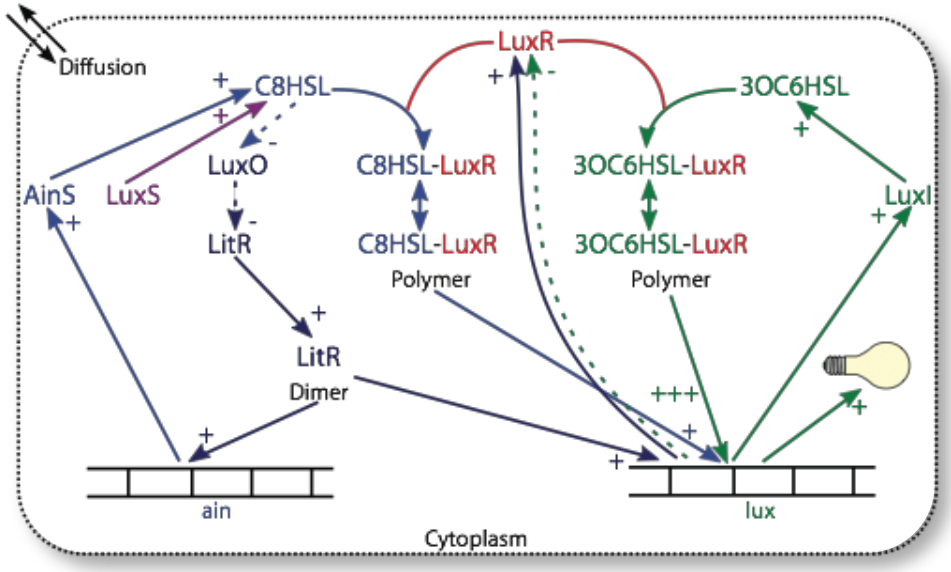
The quantity *P*(*i*,*j*) can then be interpreted as the transition probability of a diffusion process between cells.

Axes: eigenvectors of matrix P with largest eigenvalues.

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

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**

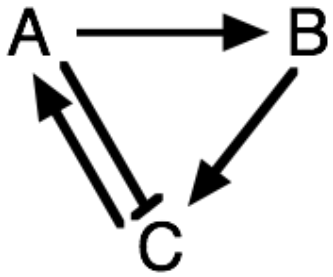
→ all states leading to an attractor = **basin of attraction**

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

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 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

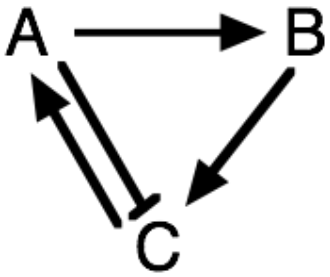
C_{i+1}	A_i	B_i
0	0	0
1	0	1
0	1	0
0	1	1

#	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**

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

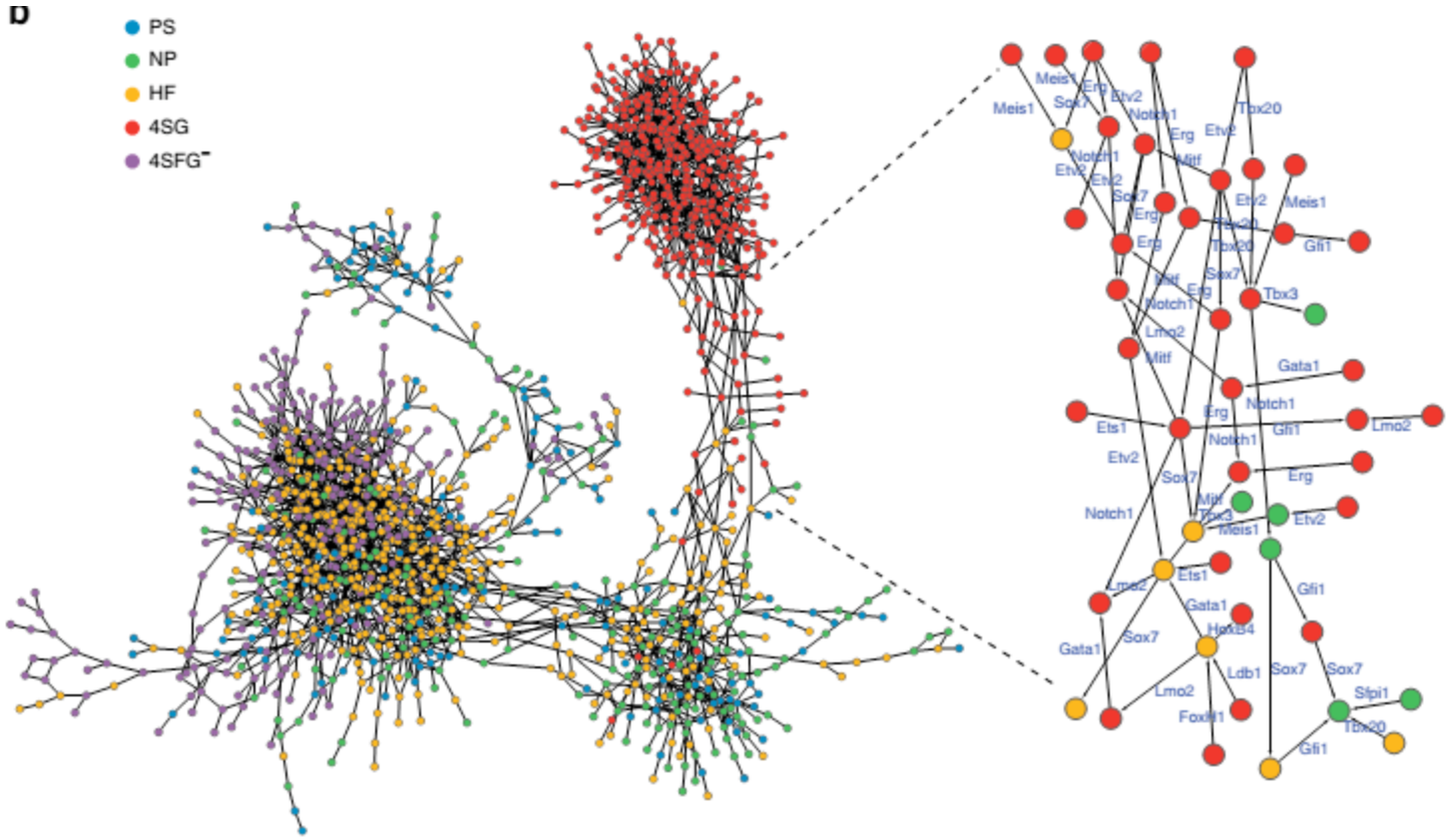
Determine suitable expression thresholds for each gene to categorize its expression levels into binary on / off states.

Note that only a small number 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)

Automatic derivation of rules for Boolean Network

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. For every variable $v_i \in V$, let N_i be the set of states without a v_i-labeled edge. For every i the number of states $s \in N_i$ such that $u_i(s) = s(i)$ is greater than or equal to t_i. (This condition "*maximizes the number of states 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)

Automatic derivation of rules for Boolean Network

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 and/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

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

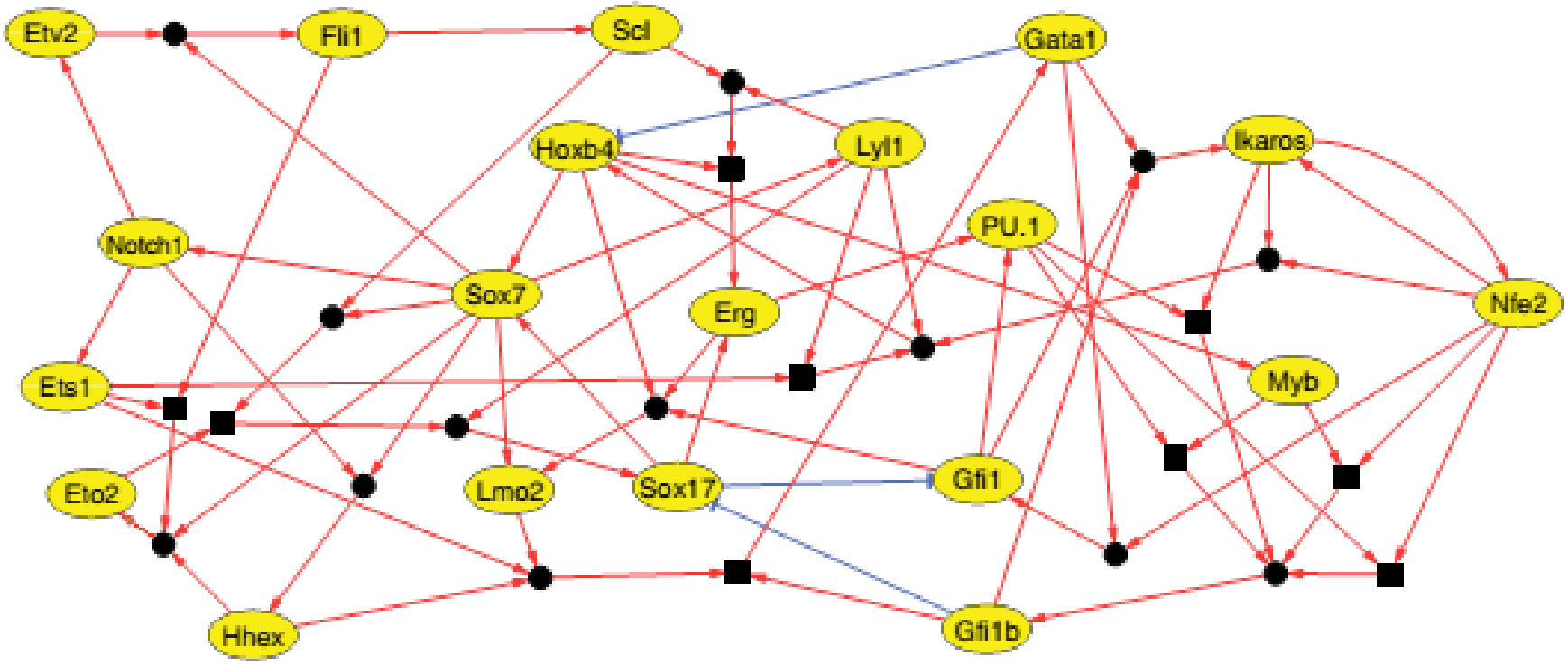
Additional validity check of the postulated rules:

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

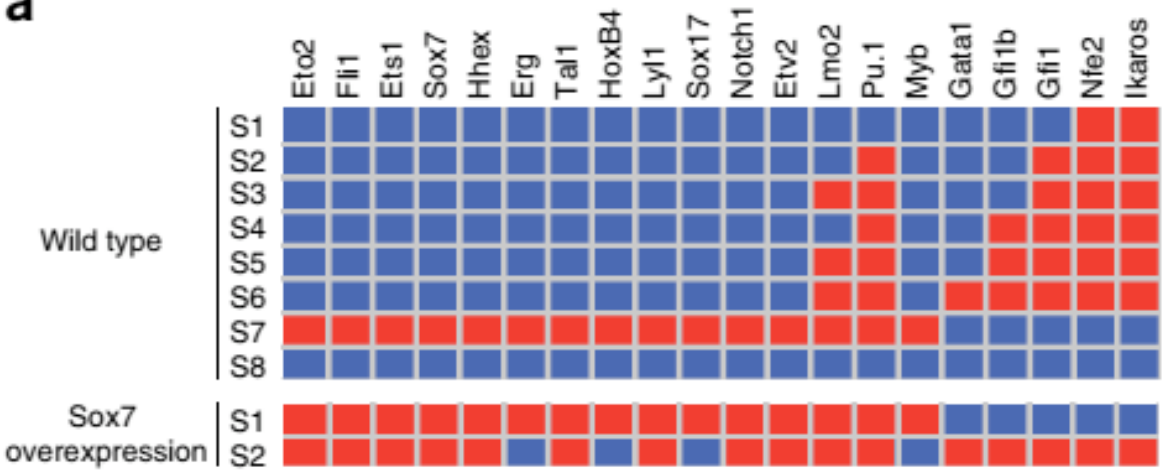
Moignard et al.,
Nature Biotech.
33, 269 (2015)

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.

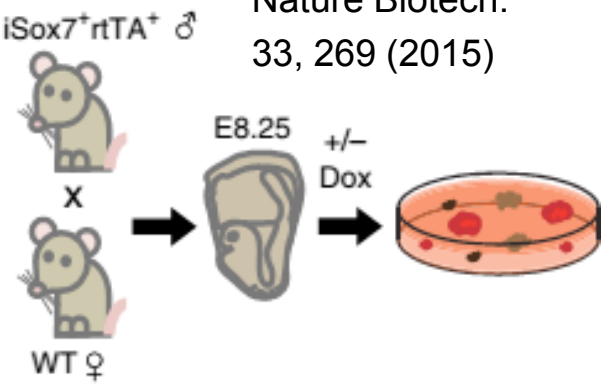
Sox7 is predicted to regulate more targets than any other TF, suggesting that perturbing its expression could have important downstream consequences

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

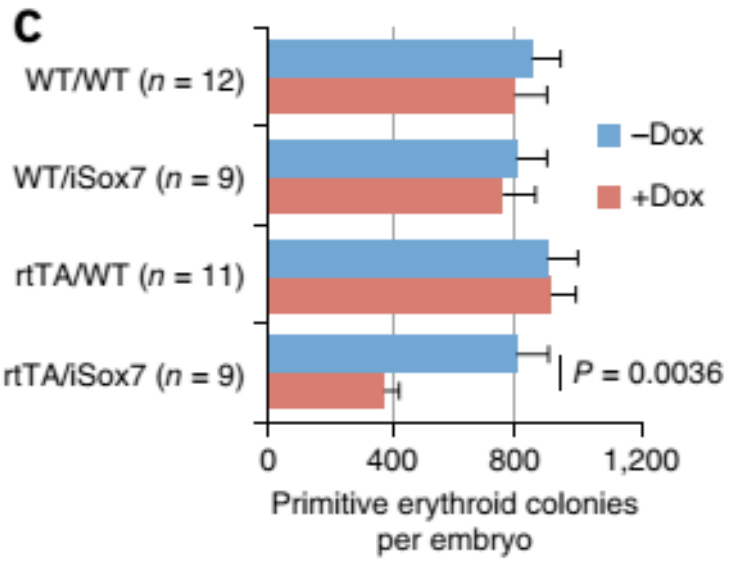
Control experiments

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.

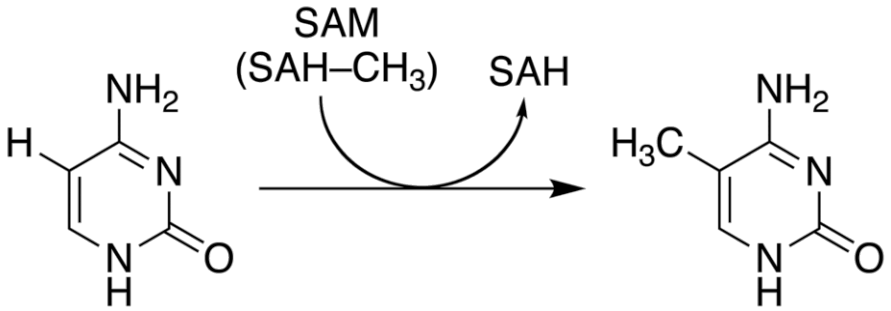
Using an automated Boolean Network synthesis toolkit we identified a core network of 20 highly connected TFs, which could reach 8 stable states representing blood and endothelium.

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

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

Cytosine methylation

Observation: 3-6 % of all cytosines are methylated in human DNA.
This methylation occurs (almost) exclusively when cytosine is followed by a guanine base -> **CpG dinucleotide**.

Cytosine 

5-methyl-cytosine

Mammalian genomes contain much fewer (only 20-25 %)
of the CpG dinucleotide than is expected by the G+C content
(we expect 1/16 ≈ 6% for any random dinucleotide).

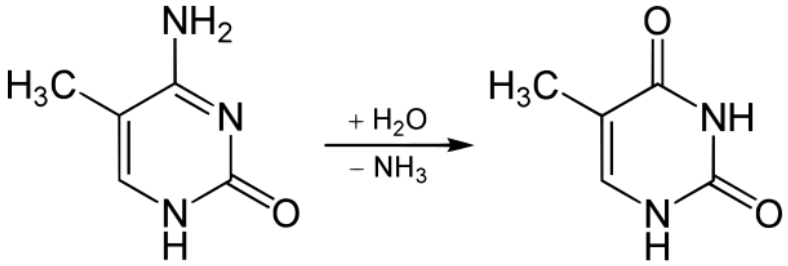
This is typically explained in the following way:

As most CpGs serve as targets of DNA methyltransferases,
they are usually methylated.

Esteller, Nat. Rev. Gen. 8, 286 (2007)
www.wikipedia.org

Cytosine methylation

5-Methylcytosine can easily **deaminate** to **thymine**.

5-methyl-cytosine 

thymine

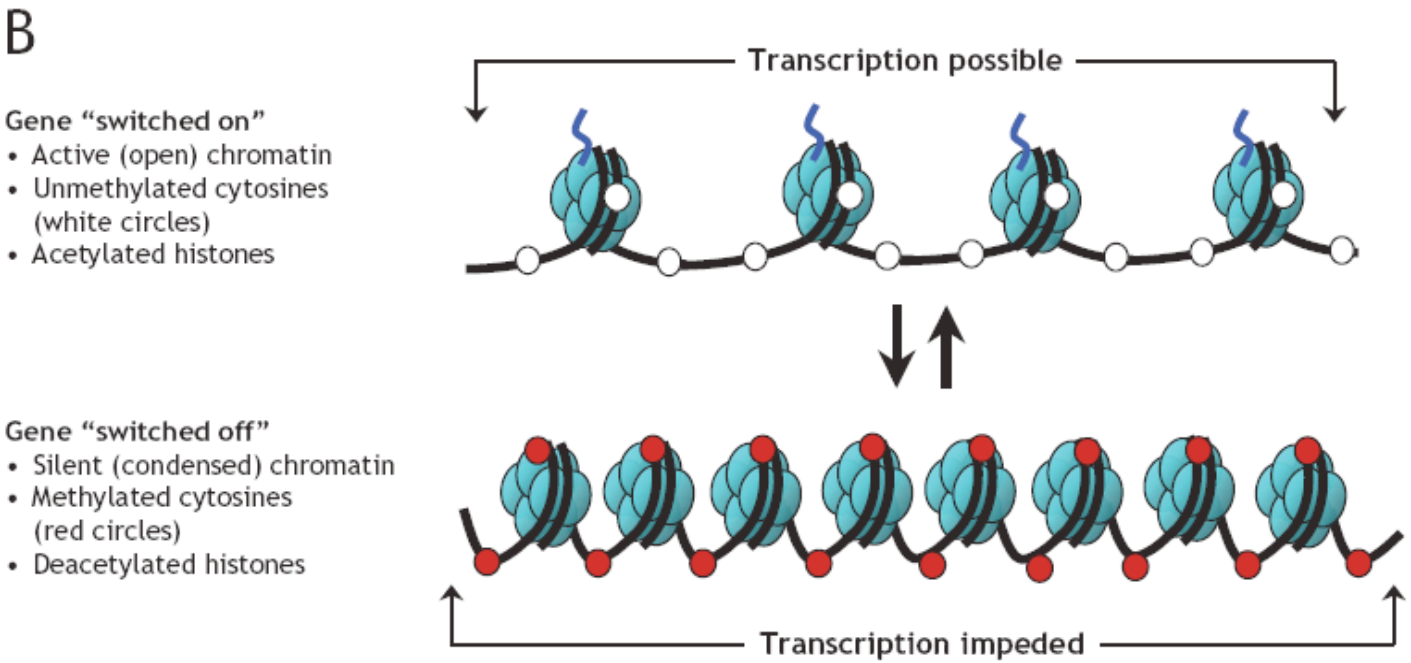
If this mutation is not repaired, the affected CpG is permanently converted to TpG (or CpA if the transition occurs on the reverse DNA strand).

Hence, methylCpGs represent **mutational hot spots** in the genome.
If such mutations occur in the germ line, they become heritable.

A constant loss of CpGs over thousands of generations
can explain the low frequency of this
special dinucleotide in the genomes of human and mouse.

Esteller, Nat. Rev. Gen. 8, 286 (2007)
www.wikipedia.org

effects in chromatin organization affect gene expression



Schematic of the reversible changes in chromatin organization that influence gene expression:

genes are expressed (switched on) when the chromatin is **open** (active), and they are inactivated (switched off) when the chromatin is **condensed** (silent).

White circles = unmethylated cytosines;

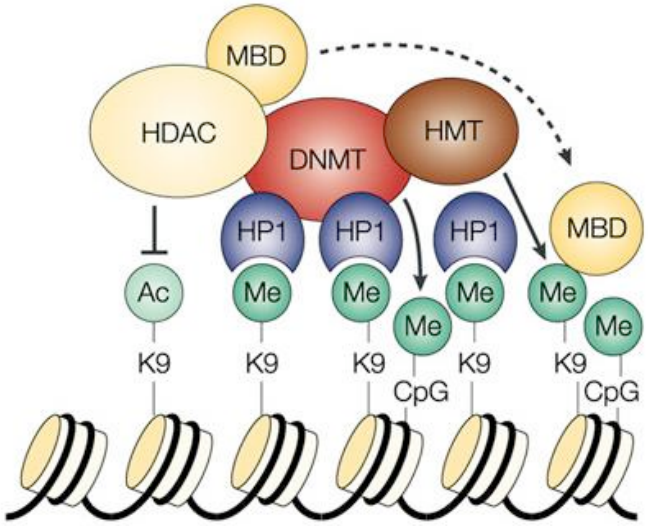
red circles = methylated cytosines.

Rodenhiser, Mann, CMAJ 174, 341 (2006)

Enzymes that control DNA methylation and histone modfications

These dynamic chromatin states are controlled by reversible epigenetic patterns of **DNA methylation** and **histone modifications**.

Enzymes involved in this process include
- DNA methyltransferases (DNMTs),
- histone deacetylases (HDACs),
- histone acetylases,
- histone methyltransferases and the
- methyl-binding domain protein MECP2.



For example, **repetitive** genomic sequences (e.g. human endogenous retroviral sequences = HERVs) are **heavily methylated**, which means transcriptionally silenced.

Rodenhiser, Mann, CMAJ 174, 341 (2006)
Feinberg AP & Tycko P (2004) Nature Reviews: 143-153

DNA methylation

Typically, unmethylated clusters of CpG pairs are located in **tissue-specific genes** and in essential **housekeeping genes.**

(House-keeping genes are involved in routine maintenance roles and are expressed in most tissues.)

These clusters, or **CpG islands**, are targets for proteins that bind to unmethylated CpGs and initiate gene transcription.

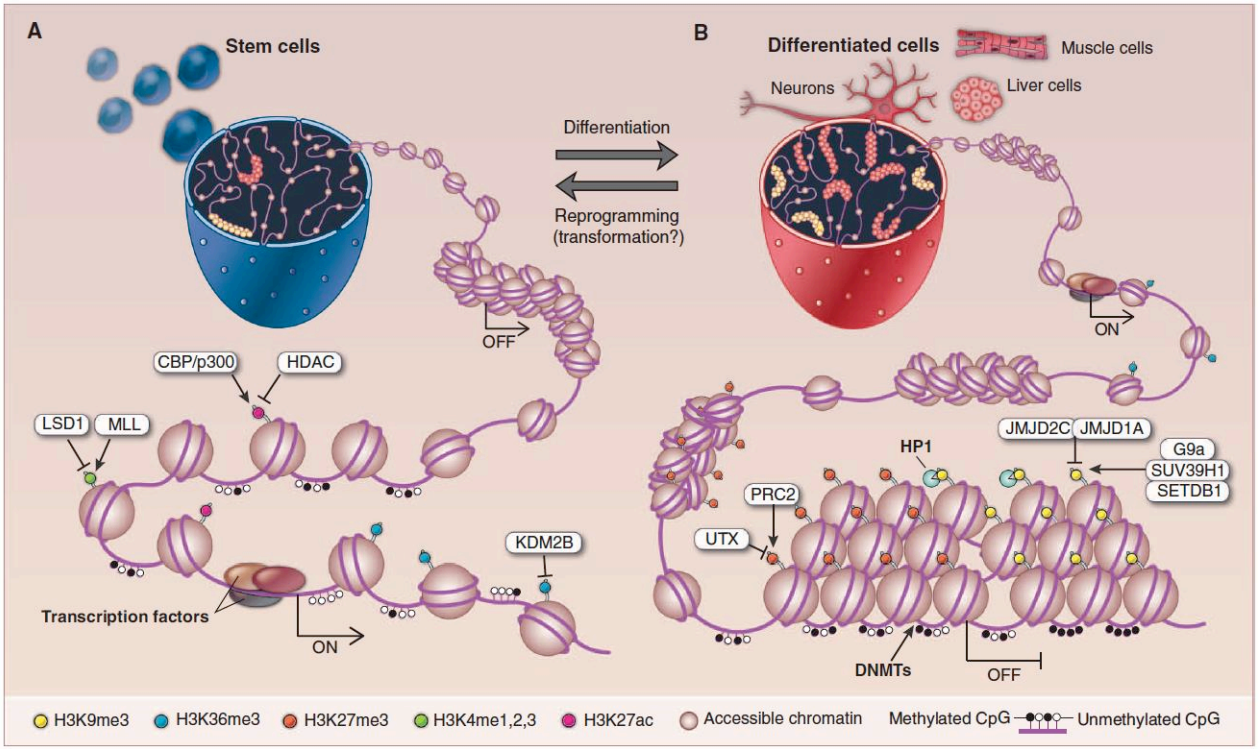
In contrast, **methylated CpGs** are generally associated with silent DNA, can block methylation-sensitive proteins and can be easily mutated.

The loss of normal DNA methylation patterns is the best understood epigenetic cause of disease.

In animal experiments, the removal of genes that encode DNMTs is lethal; in humans, overexpression of these enzymes has been linked to a variety of cancers.

Rodenhiser, Mann, CMAJ 174, 341 (2006)

Differentiation linked to alterations of chromatin structure



(B) Upon differentiation, inactive genomic regions may be sequestered by repressive chromatin enriched for characteristic histone modifications.

(A) In pluripotent cells, chromatin is hyperdynamic and globally accessible.

ML Suva et al. Science 2013; 339:1567-1570

Epigenetic stability

In somatic tissues, **CpG islands** at housekeeping or developmental promoters are largely unmethylated, whereas non-regulatory CpGs distributed elsewhere in the genome are largely methylated.

This **DNA methylation landscape** is relatively **static** across all somatic tissues.

Most of methylated CpGs are pre-established and inherited through cell division.

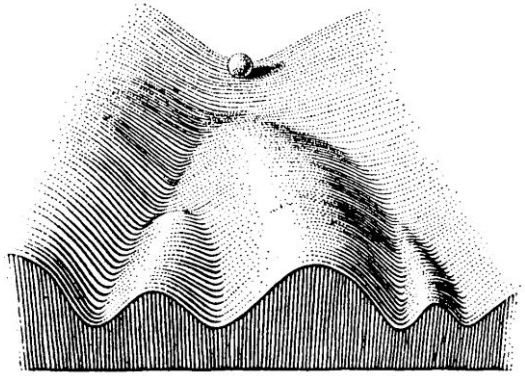
In at least two phases of the life cycle of mammals, epigenetic stability is globally perturbed:

- when gametes fuse to form the zygote and
- when gamete precursors (primordial germ cells; PGCs) develop and migrate in the embryo.

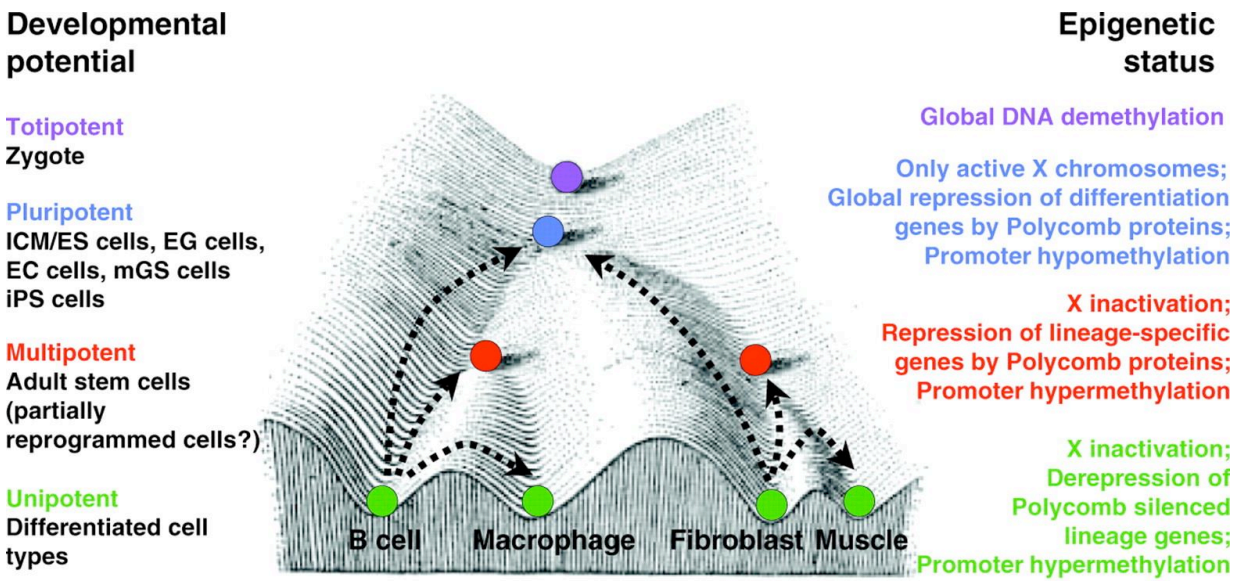
This ***in vivo*** **'reprogramming'** of the epigenetic landscape signals the reacquisition of totipotency in the zygote and the formation of the next generation through PGCs.

Cantone & Fisher,
Nature Struct Mol
Biol. 20, 292 (2013)

Waddington: Epigenetic landscape

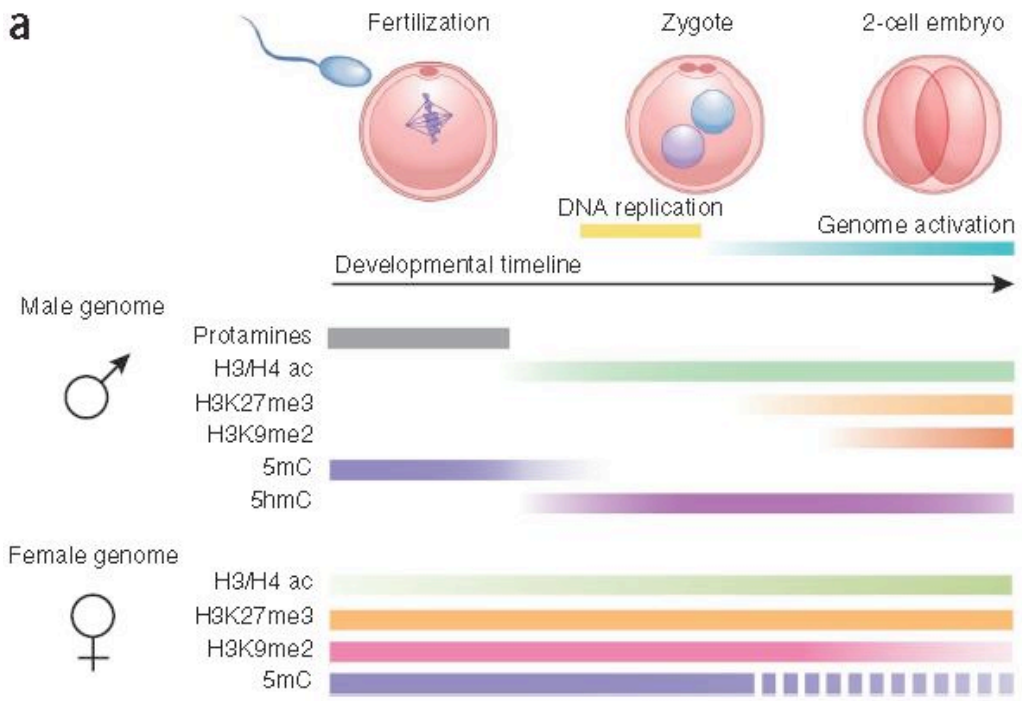


Conrad H. Waddington 1956: "Principles of Embryology"; www.nature.com



Konrad Hochedlinger and Kathrin Plath,
Development 136, 509-523 (2009)

Epigenetic changes during in vivo reprogramming



Global DNA and histone modifications that lead to transcriptional activation of the embryonic genome between the late zygote (paternal genome only) and the 2-cell stage.

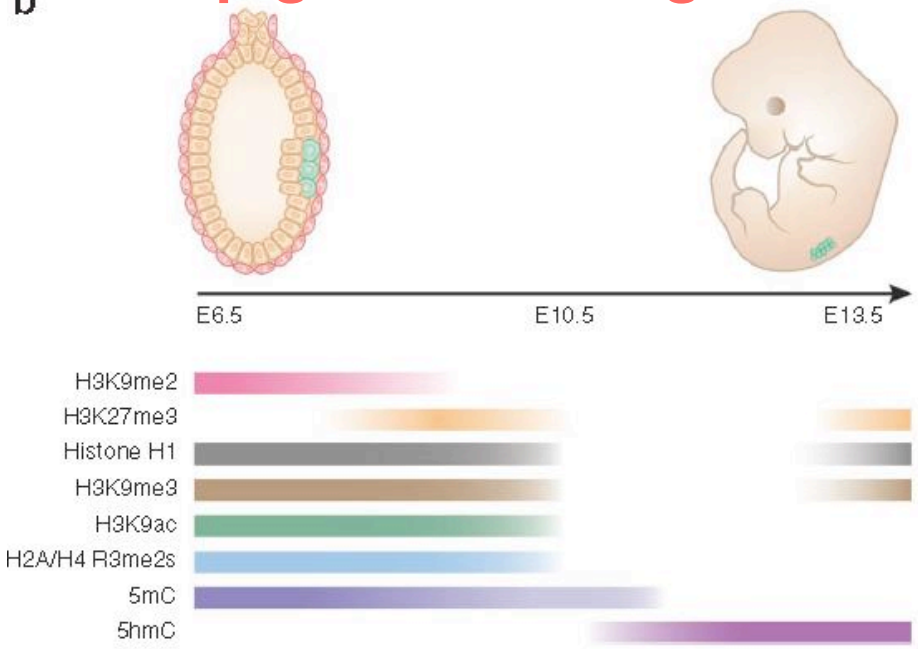
Protamines are small, arginine-rich, nuclear proteins that replace histones late in the haploid phase of spermatogenesis and are believed essential for sperm head condensation and DNA stabilization.
In humans, 10-15% of the sperm's genome is packaged by histones thought to bind genes that are essential for early embryonic development (www.wikipedia.org).

Gamete genomes undergo different epigenetic programs after fertilization.

The **paternal** genome is mostly subject to epigenetic remodeling at the zygote stage. The **maternal** genome gradually loses repressive modifications during the subsequent cleavage divisions.

Cantone & Fisher, Nature Struct Mol Biol. 20, 292 (2013)

Epigenetic changes during germline development



Global epigenetic changes during germline development from PGC specification (E6.5) to the mitotic/meiotic arrest at E13.5.

Two major reprogramming phases can be distinguished during PGC migration toward the genital ridges (E7.5–E10.5) and upon their arrival into the gonads (E10.5–E12.5).

Cantone & Fisher, Nature Struct Mol Biol. 20, 292 (2013)