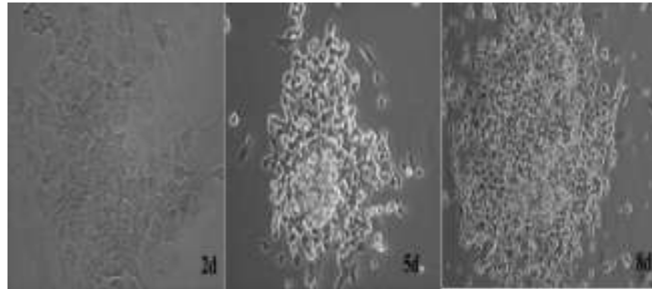
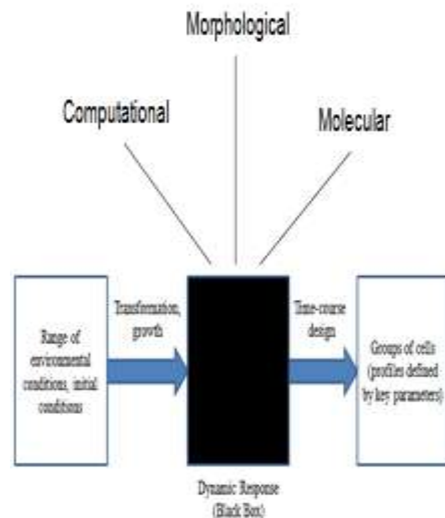
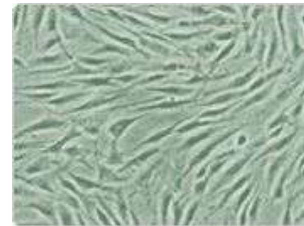
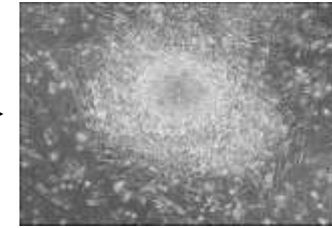
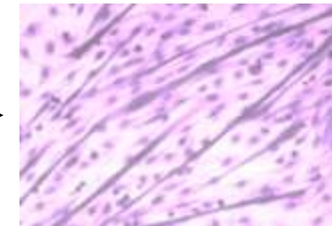
Biocomplexity of Inducible Cells

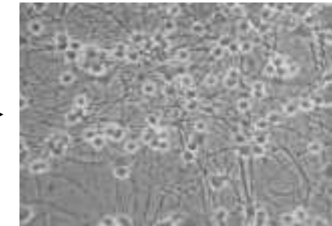












Bradly Alicea

Michigan State University

http://www.msu.edu/~aliceabr

Synthetic Daisies

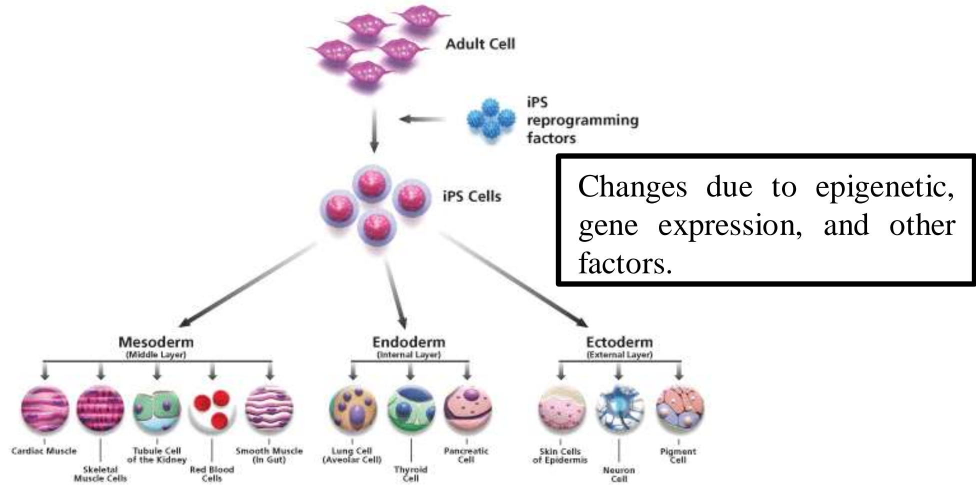
http://syntheticdaisies.blogspot.com

Presented to the Embryo Physics Course, Silver Bog, Second Life

http://www.embryophysics.org

Q: What are “inducible” cells?

Modify cells for purposes of production, studying disease, or transplantation. Cartoon shows animal cells, but *E.coli* can also be used (e.g. synthetic biology).

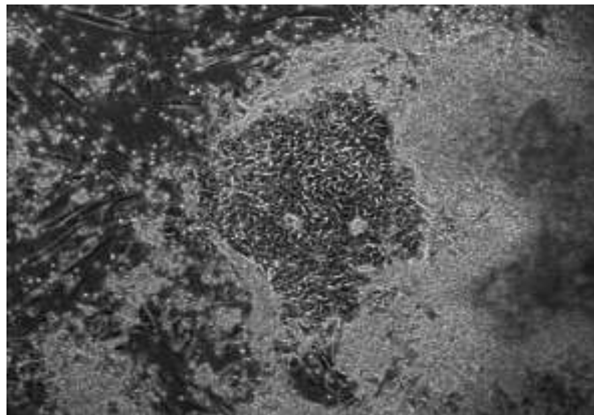


Changes due to epigenetic, gene expression, and other factors.

COURTESY: Sigma-Aldrich

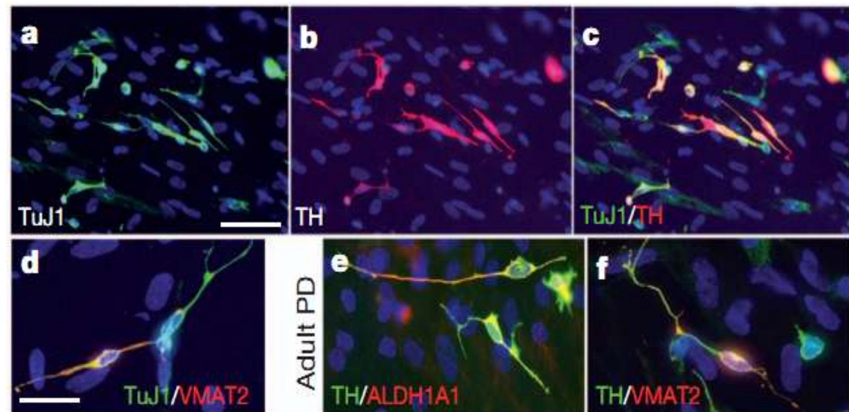
A: Cells that can be transformed from one phenotype to another.

Thomson Lab, University of Wisconsin



iPS cells (pluripotent)

Caiazzo, M. et.al Nature, 476, 224-227 (2011).



iN cells (neuronal)

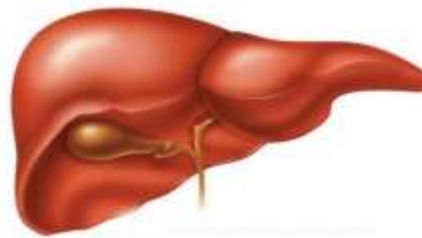
Q: What is biocomplexity?

Non-reductionist approach, incorporates multiple mechanisms and levels of organization.

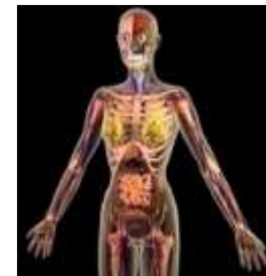
COURTESY: NSF, Wikipedia

A: complexity of function and structure in living systems.

Eukaryotic Cells



Organs (cell populations)



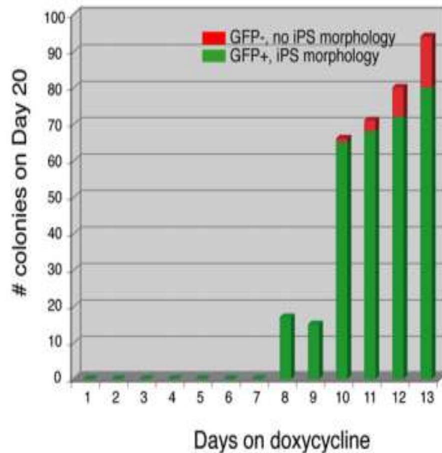
Organisms

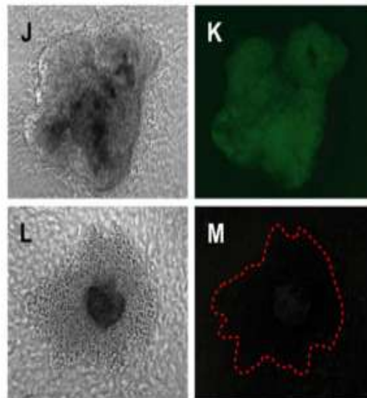
COURTESY: Stem Cell School (http://stemcellschool.com/)

Q: How do we “induce” animal cells to new phenotypes?

A: Polycistronic viral vector (e.g. OK-SIM, BAM) that encode transcription factors (for key trigger genes).

Doxycyclene treatment: shuts down factors (result: GFP $^{-}$ colonies).





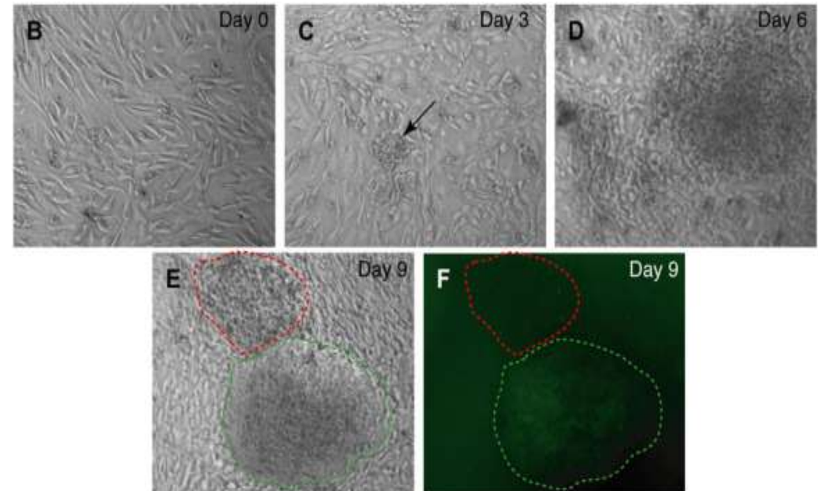
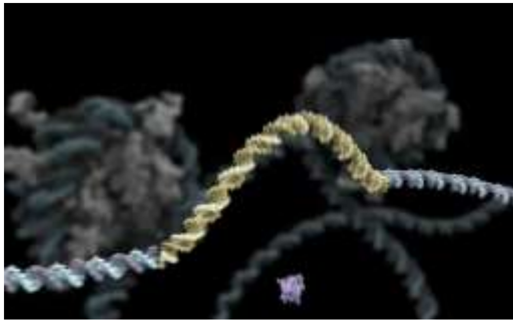
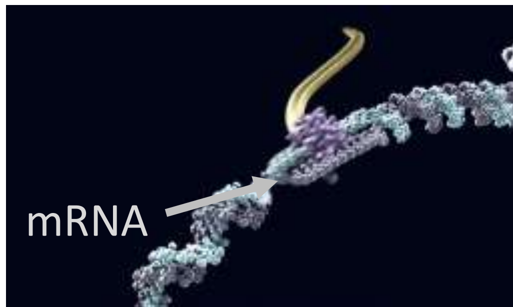


Figure 1, Stadfeld, M. et.al, Cell Stem Cell, 2, 230-240, (2008).

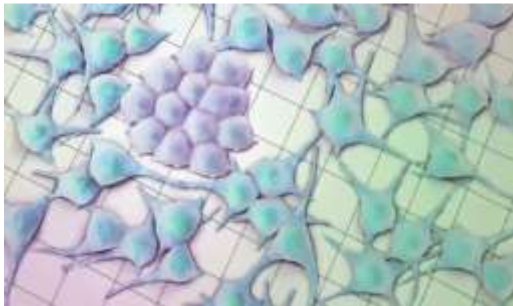
“Four factor” approach of Yamanaka (2007).



Retroviral vector (delivered via transgene) integrates into genome of “infected” cells. Occurs in only a fraction of exposed cells.



Induces differential gene expression, leads to changes in mRNA concentrations, protein production, cell phenotype.



Colonies form among cells with the new “pluripotent” phenotype, (aggregation of pluripotent cells).

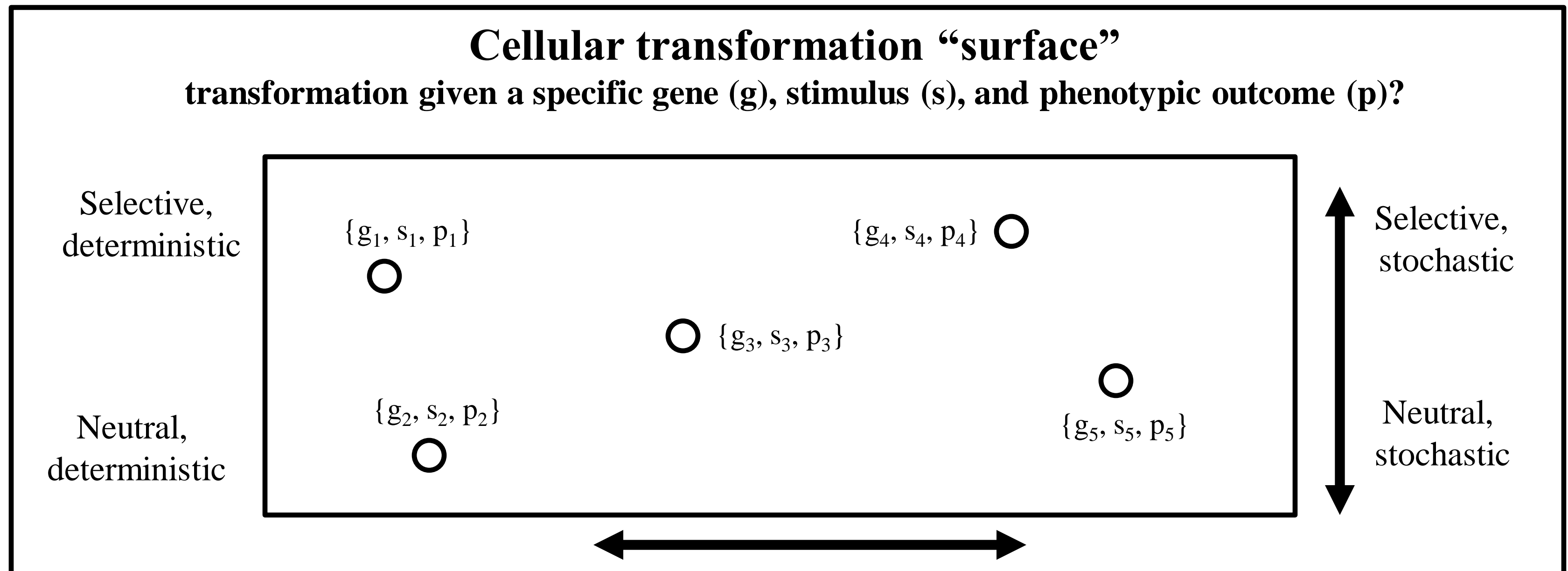
COURTESY: Stem Cell School (http://stemcellschool.com/)

Dynamic transformation: Observing the change in a cellular phenotype over time (direct) and/or across diversity (indirect).

CONVENTIONAL VIEW: state as a specific, predetermined (*a priori*) phenotype and biochemical indicators.

NEW VIEW: transformation is a position on a "surface", governed by 3 parameters. Decomposed using experiment, computation.

Adapted from Yamanaka (2009), Hanna et.al (2009).



Attractor basin approach:

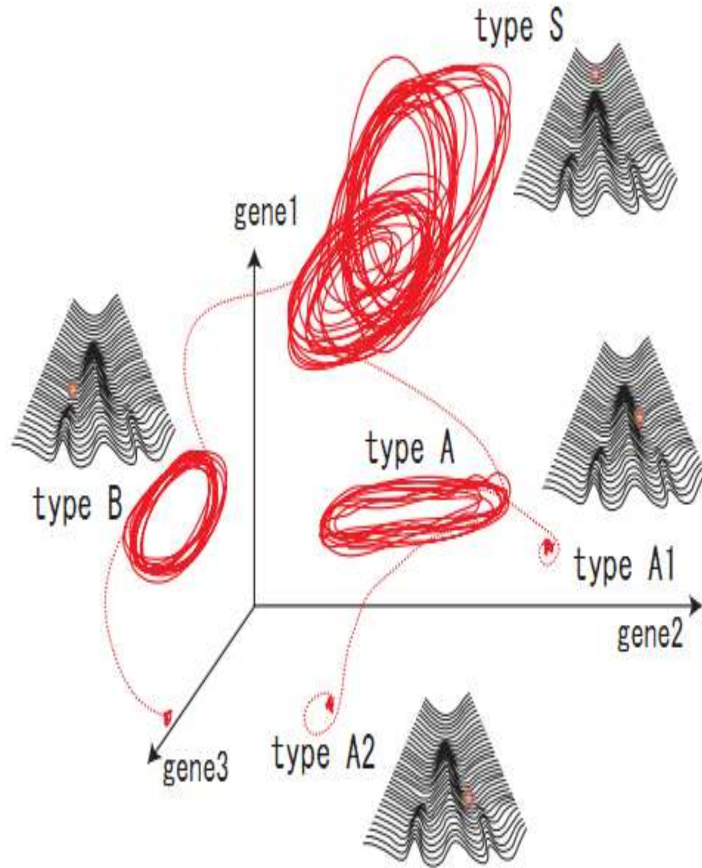


Figure 1

A Schematic representation of phase-space which represents expression dynamics of several cell types. Each axis shows the expression level of a gene. The trajectory in the phase space represents the time course of expression profile, where attractors correspond to distinct cell types, i.e., type S, A1, A2, and B. In this example, type-S cells act as stem-type cells, which can differentiate into type-A, B, while differentiated types lose the potential to differentiate into other cell types. The epigenetic landscape of cellular states is also shown, in which the change of cellular state is represented as the trajectory of a ball falling along branching valleys.

Figure 1, Furusawa and Kaneko, Biology Direct, 4: 17 (2009).

* pluripotent phenotypes are in "metastable states" (can be nudged into multiple attractor basins given stimulus).

* differentiated phenotypes can be in attractor basins, or can be metastable themselves.

* consistent with Waddington's model of developmental canalization (specific pathways for differentiation).

* cells choose a specific "path" to and from different cellular states.

Energy landscape used to characterized transitional and end states during reprogramming.

* left: navigation along minima during process (characterize molecular changes).

* right: stable states = minima and maxima. Characterize cells in relation to differentiated and stem-like cell type.

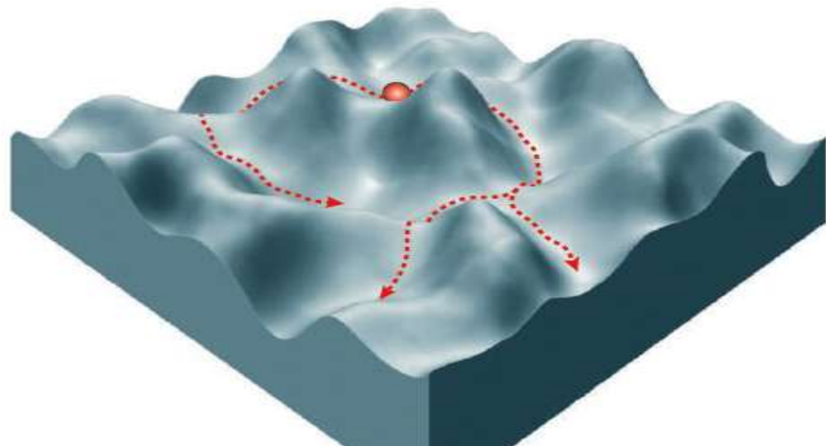
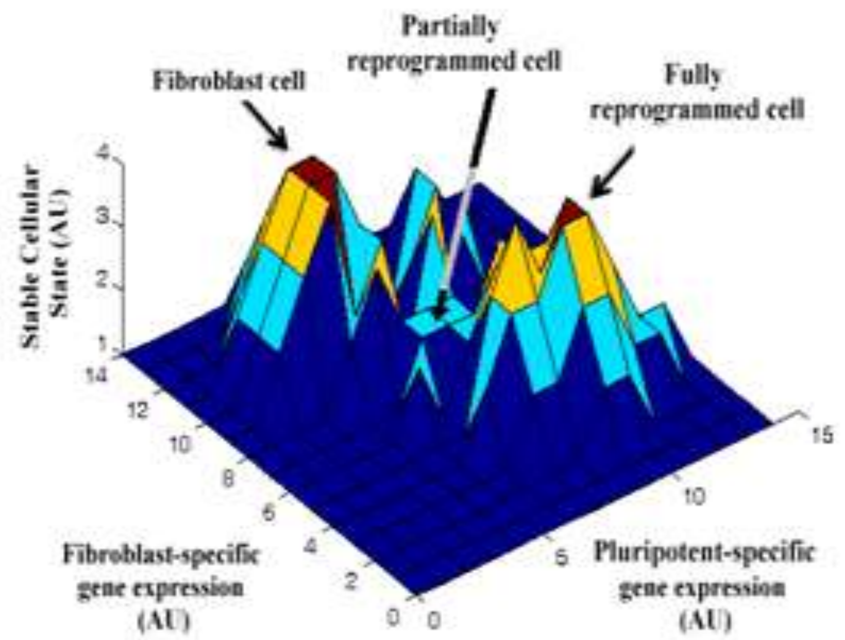


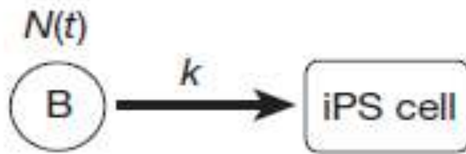
Figure 2 | **Cellular reprogramming as navigation through a complex attractor landscape.** In a complex cellular attractor landscape there might be many coexisting stationary attractors (here represented as local minima), each of which might be associated with a unique molecular signature. In this view, cellular reprogramming corresponds to guiding the cell through the landscape from one local minimum to another (shown by the dotted arrows). As there might be many distinct paths between minima (both direct and through intermediary minima), reprogramming from one cell type to another might be achieved though numerous different routes[5,108,120].

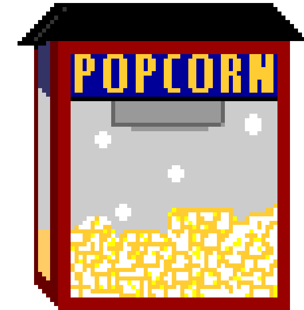
MacArthur et.al Nature Reviews Molecular Cell Biology, 10, 673 (2009).



Three components to the stochastic model of reprogramming (presented in Hanna et.al, 2009):

1) basic assumption: given N cells, one-step reprogramming process occurs with constant cell-intrinsic rate k. "Popcorn" metaphor.

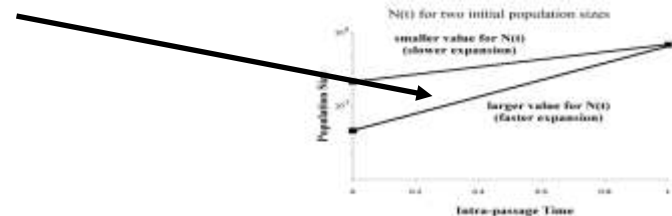


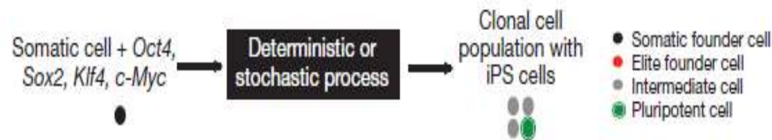


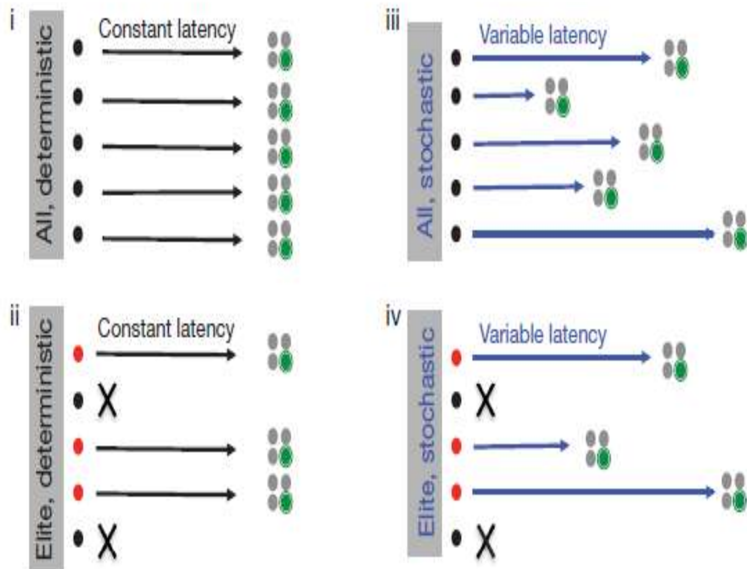
2) latency: interval t_p defined as time between t_n (when first cell in population N is reprogrammed) and t_{n+1} when daughter cells grow to reach detection threshold. From "first settler" to viable colony.

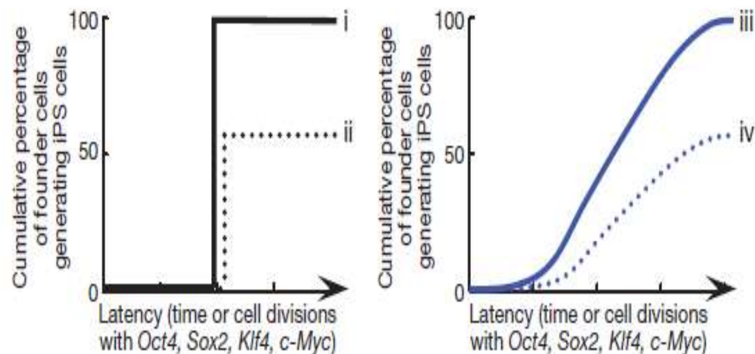
3) scaling: at any t_n, population of cells in a well, $N(t)$, scales at rate which first reprogramming event takes place (determines slope). Cumulative PDF:

$$P(t + t_p) \approx 1 - e^{-k\tau} \text{ where } \tau = \int_0^t N(t')dt'$$
$$\approx 1 - e^{-kN_{eff}(t-t_0)} \text{ for } t > t_0$$









Hanna et.al, Nature 462, 595-601 (2009).
Yamanaka, Nature, 460(7251), 49-52 (2009).

Competing models for reprogramming (stochastic vs. deterministic):

1) stochastic: transformation occurs according to a **variable** latency.

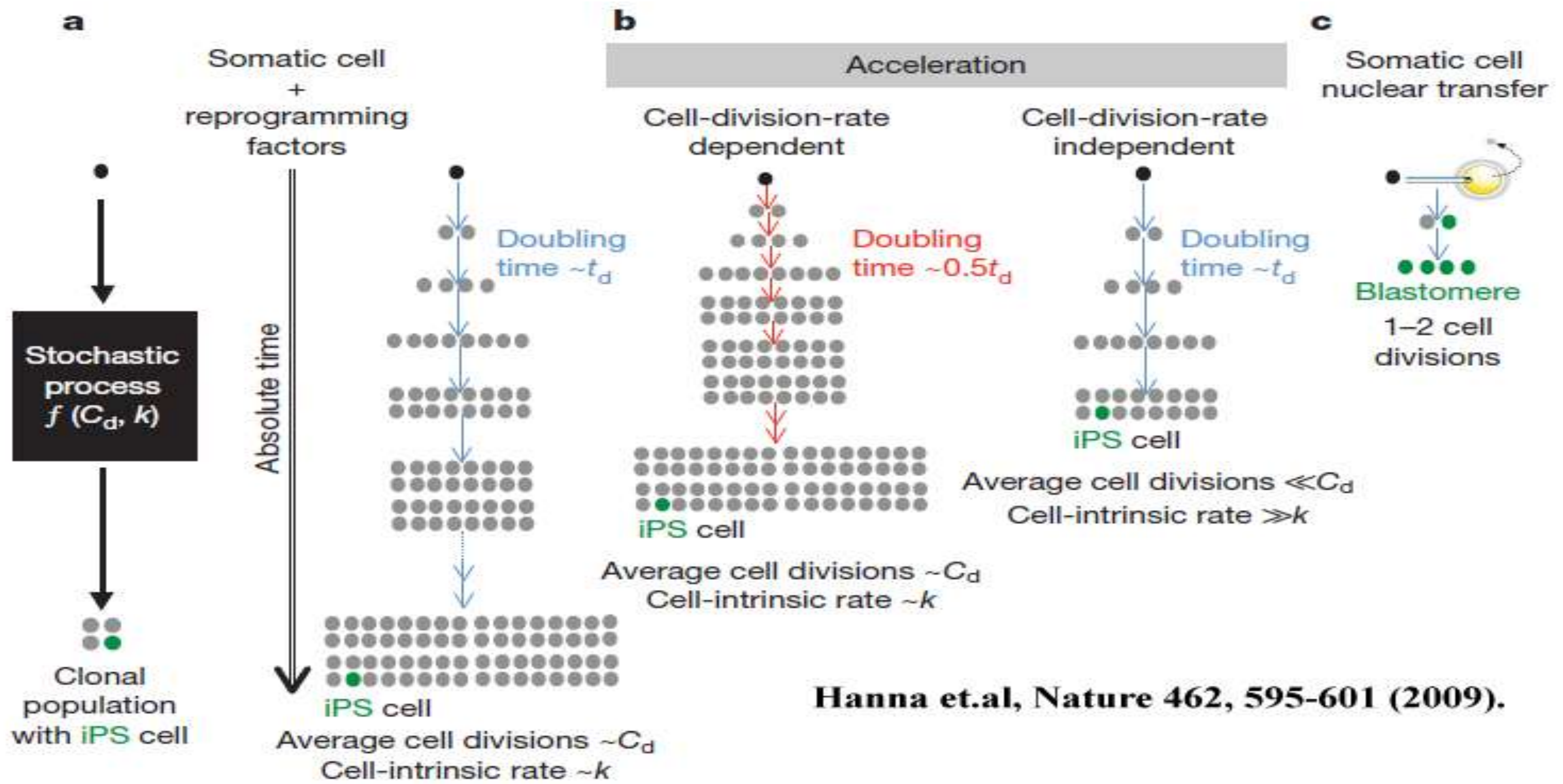
* time from trigger to transformation is variable (cell cycle $c = m$ transformations).

2) deterministic: transformation occurs according to a **uniform** latency.

* time from trigger to transformation is uniform.

Elite models argue that only a subset (*1/n*) of cells will reprogram (innate ability).

* elite model is independent of stochasticity vs. uniformity (independent mechanisms?)



Example of scaling (more instances of transformation with more cells):

$N = 10^3$ time to reach > 90% reprogrammed cells in well longer.

$N = 10^6$ time to reach > 90% reprogrammed cells in well shorter.

Simplest scenario (for stochastic reprogramming):

* one-step rate-limiting transition characterized by a cell-intrinsic rate, which does not describe reprogramming behavior before and soon after transgene induction.

* perhaps there are multiple modes of reprogramming acceleration.

Arrive at this model by considering:

1) closely monitoring transgene induction, 2) plating efficiency, 3) cell proliferation, 4) changes in population size across experiment.

There is a yet-to-be-defined rate-limiting, continuous stochastic mechanism (according to model):

* function of cell division before fully reprogrammed state is attained.

* results support "all with variable latency" model (neither "elite" nor "deterministic").

* might iPS cells arise preferentially from a precursor (progenitor or adult SC)?

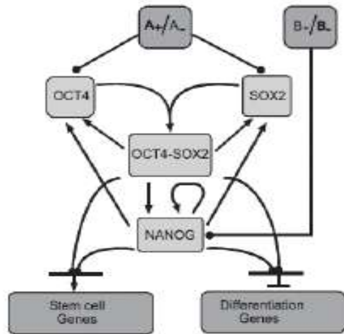


Figure 1. The Core Transcriptional Network of the ES Cell
Genes and proteins are represented as single entities. Each outgoing arrow represents a protein (the outgoing merging arrows from OCT4 and SOX2 represent complex formation). Each ingoing arrow represents a protein with a role as a TF. Signals A_{+}, e.g., the wnt signaling protein, triggers the system to sustain self-renewal and pluripotency, whereas signals B_{-}, e.g., p53, shuts it off, thereby leading to differentiation. It should be noted that there are also signals A_{-} that repress OCT4 and SOX2. These variants are not shown here since the effects will be similar to those of B_{-}. OCT4, SOX2, and NANOG individually as well as jointly target stem cell and differentiation genes. We model the set of TGs that are jointly regulated by OCT4, SOX2, and NANOG. The nature of the regulation at the TGs is not specified, since we explore all the possibilities. However, the final effect of the regulation is indicated.
DOI: 10.1371/journal.pcbi.0020123.g001

Figure from PLoS Computational Biology, 2(9), e123 (2006).

Hypothesis: Cellular reprogramming can be driven by noise.

* noise in the form of transcriptional variance and other stochastic processes can trigger, drive reprogramming process *in vitro*.

* presence of Oct4, Sox2, and NANOG suppress differentiation genes and activate stem cell genes (modules).

Example: MacArthur et.al, PLoS One, 3(8), e3086.

Computational approaches to gene expression include adding noise (stochastic element) to model.

* non-specific noise in expression of four factors, other genes can trigger reprogramming.

Black function: Oct4, Sox2. Blue function: NANOG. Red function: lineage-specific master genes, σ: parameter value for amplitude of noise (same for every gene).

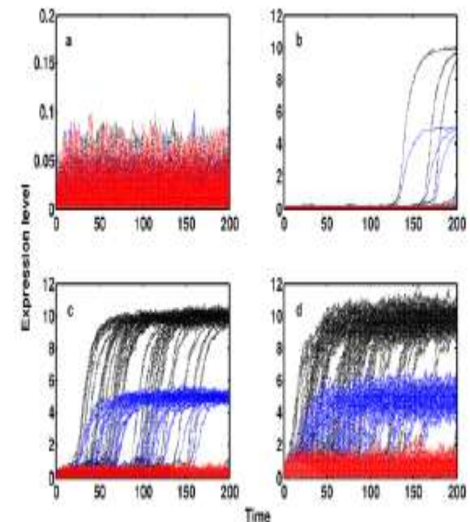


Figure 5. Non-specific noise can trigger reprogramming to a pluripotent state.

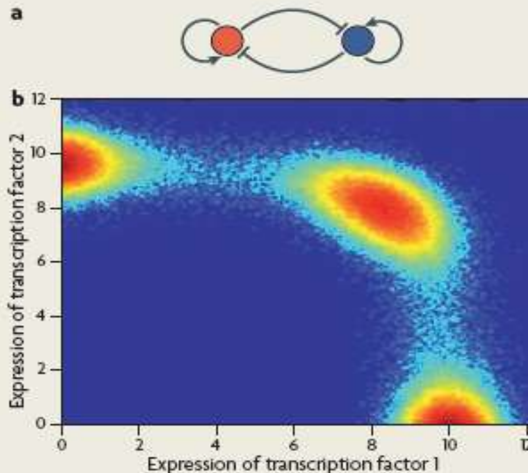
Figure from PLoS One, 3(8), e3086 (2008).

Box 4 | **A stochastic multi-stable switch**

Consider the simple motif in which two transcription factors activate their own expression and mutually repress each others' expression (see the figure, part a). This type of feedback naturally gives rise to multi-stability[86,126] and provides the cell with the ability to make all-or-none fate decisions in response to external cues. The following stochastic differential equations describe the expression levels of two transcription factors (x_1 and x_2) that are interacting in this way:

$$dx_1 = \frac{x_1^2}{K_1 + x_1^2 + k_1 x_2^2} - b_1 x_1 + \sigma_1 dW \qquad dx_2 = \frac{k_2 x_2^2}{K_2 + k_3 x_1^2 + x_2^2} - b_2 x_2 + \sigma_2 dW$$

In these equations k_1, k_2 and k_3 are the (normalized) rate constants at which transcription factors bind to promoters; K_1 and K_2 are (normalized) dissociation rate constants; b_1 and b_2 are (normalized) decay rate constants; σ_1 and σ_2 are constants determining the amplitude of noise in the system; and W denotes a Weiner process (Brownian motion). In this simple illustrative case we have assumed that each transcription factor binds cooperatively to its own promoter and to that of the other transcription factor as a homodimer (which is why x is raised to the power of two). In the absence of molecular noise ($\sigma_1=\sigma_2=0$) this model has many coexisting steady state attractors (for appropriate parameter regimes). In the presence of molecular noise ($\sigma_1,\sigma_2>0$), individual cells do not settle at a single attractor but instead stochastically switch between distinct states at a rate that depends on the amplitude of molecular noise. However, over time the joint probability density $p(x_1, x_2)$ (that is, the probability of finding a cell with expression levels of (x_1, x_2)) settles to a stationary state, and a robust distribution of cell types is achieved. The figure (part b) shows the stationary probability distribution for a representative simulation of this system: red hot spots indicate preferred genetic configurations at which cells will accumulate, and blue indicates low probability configurations.



MacArthur et.al Nat Rev Cell Biology, 10, 673 (2009).

There are many potential outcomes of reprogramming (iPS, piPS):

Stemness = what do the diversity of induced stem cells types have in common?

* pluripotency, gene regulation profiles.

* multi-stability (ability to change state in response to environmental, viral cues).

Switch that governs this transformation may be stochastic:

* Two factors activate their own expression, mutually repress each other (all-or-nothing response).

* Weiner process (additive) = stochasticity. At $\sigma = 0$, switch between fate at rate r.

OK-SIM reprogramming amounts to "blind refactoring":

* blind refactoring = systematically rewriting a program, but also affecting other parts of the program without appropriate control.

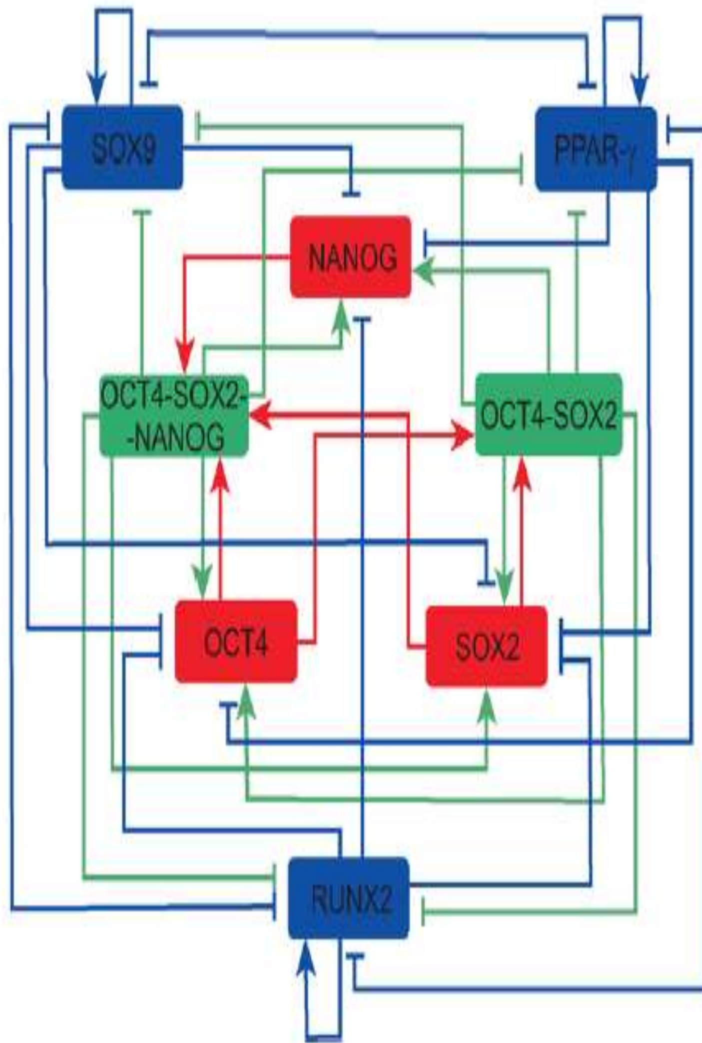
* efficacy of OK-SIM approach: ranges from 0.0001% to 29%. 290,000-fold difference! What causes this amount of variation?

A: rare pathways. Thousands of potential pathways (even in same culture), how much does each pathway contribute to overall number of cells reprogrammed? Some are common modes of action, others more itinerant.

Computational model of cellular identity:

* not possible to build a programming model (or even a mechanistic description) of reprogramming process.

* next best thing: approximate the general architecture – transcription factors added, results in a hierarchical regulatory cascade.



MacArthur et.al, PLoS One, 3(8), e3086.

Regulatory interactions among genes in core pluripotency module:

* Oct4-Sox2 and Oct4-Sox2-NANOG repress both Sox9 and PPAR-γ (seen in neural and bone cells, respectively).

* Sox9 and PPAR-γ are co-repressive of each other and Oct4, Sox2, NANOG.

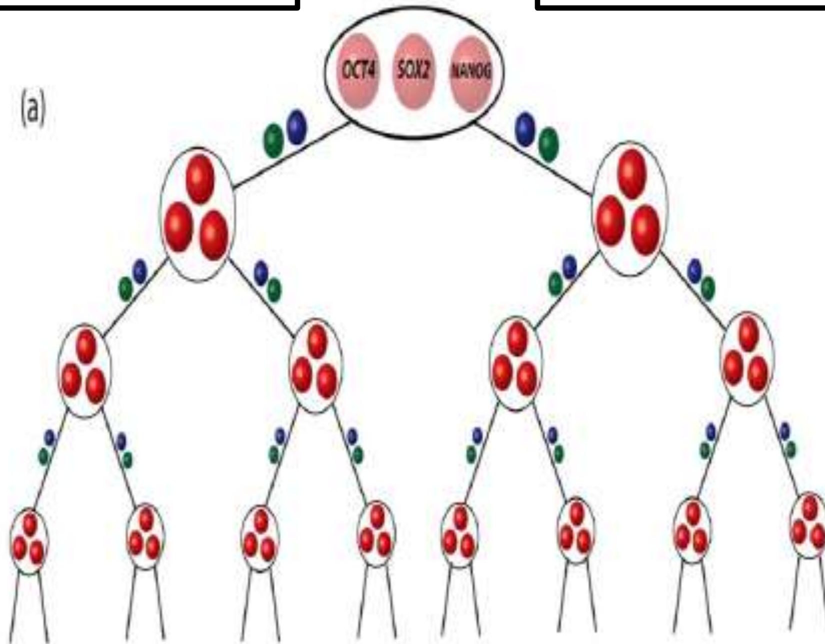
Tip of the iceberg viewpoint:

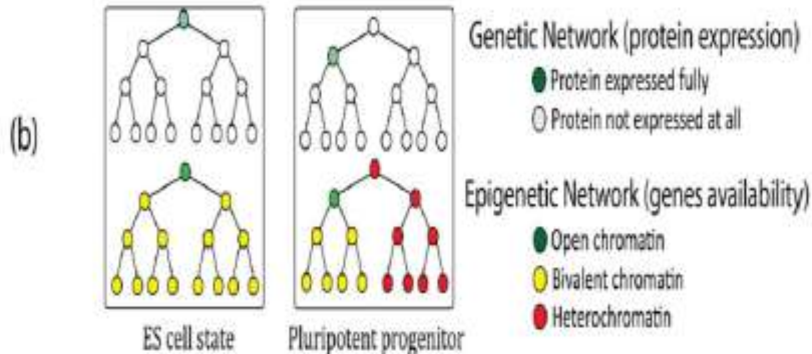
* reprogramming is a critical process that is triggered by the right state of this core module.

What role does stochasticity play? Oct4/Sox2 and NANOG vs. rest of genome.

Many possible routes to cellular state, including “rare” pathways

Master control genes govern discrete output (switch-like)





COURTESY: Artyomov et.al PLoS Computational Biology, 6(5), e1000785.

Cayley Tree model of reprogramming (described in Artyomov et.al, 2010):

Gene modules (groups of genes that work together) in a hierarchy.

* branches = cell state.

* nodes = state of epigenome, gene expression level (gene module).

* active transcription (1), not active transcription (0).

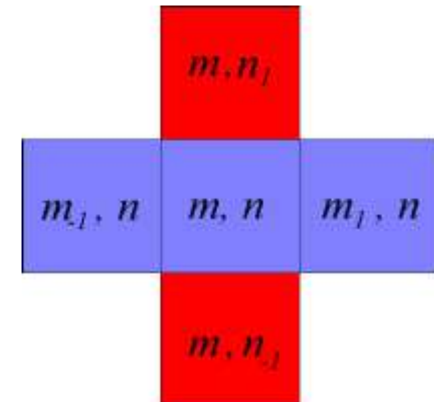
* for each trajectory, fraction of cells reprogram (e.g. “rarity” of pathway).

Protein expression (gene network-dependent) vs. gene availability (epigenetic network-dependent).

Cellular Automata: discrete dynamical simulation.

Cells have **properties** and **interaction rules**, behave in **parallel**.

Below: 2-D von Neumann neighborhood, order 1



Properties: internal state.

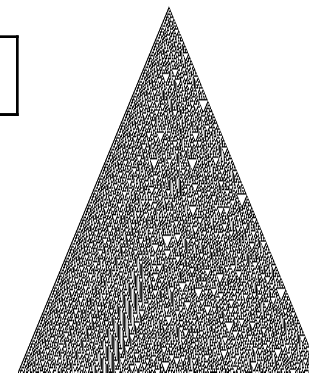
Interaction rules: if n > 2 neighbors are red, turn red.

Parallelism: all cells use same set of rules, have same properties.

Example: Wolfram's Rule 30 (1-D lattice)

1	2	3	4	5	6	7	8

current pattern	111	110	101	100	011	010	001	000
new state for center cell	0	0	0	1	1	1	1	0



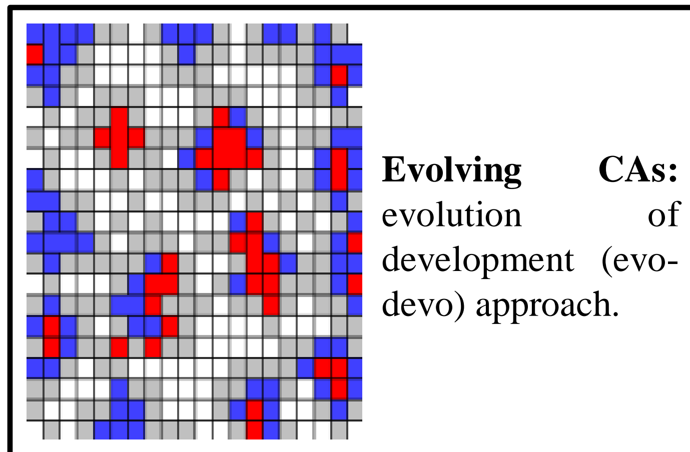
Rule 30 - model

Rule 30 - nature

How to "recapitulate" morphogenesis:
Intrinsic noise = stochastic process in parallel = order. Change in development AND evolution among hair pigment cells (see Pilkus et.al, Science, 332, 586-589 – 2011).

Wolfram's 2-D rules: self-organized patterns (development), evolve in time (evolution).

Way to understand population-level morphological changes: evolving CAs (digital evo-devo). Couple with genetic algorithms.

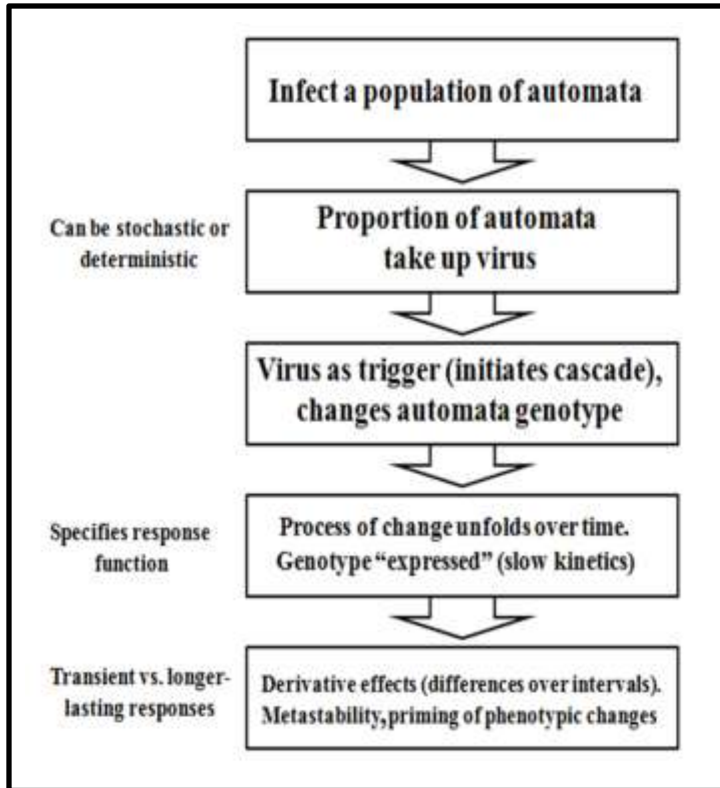


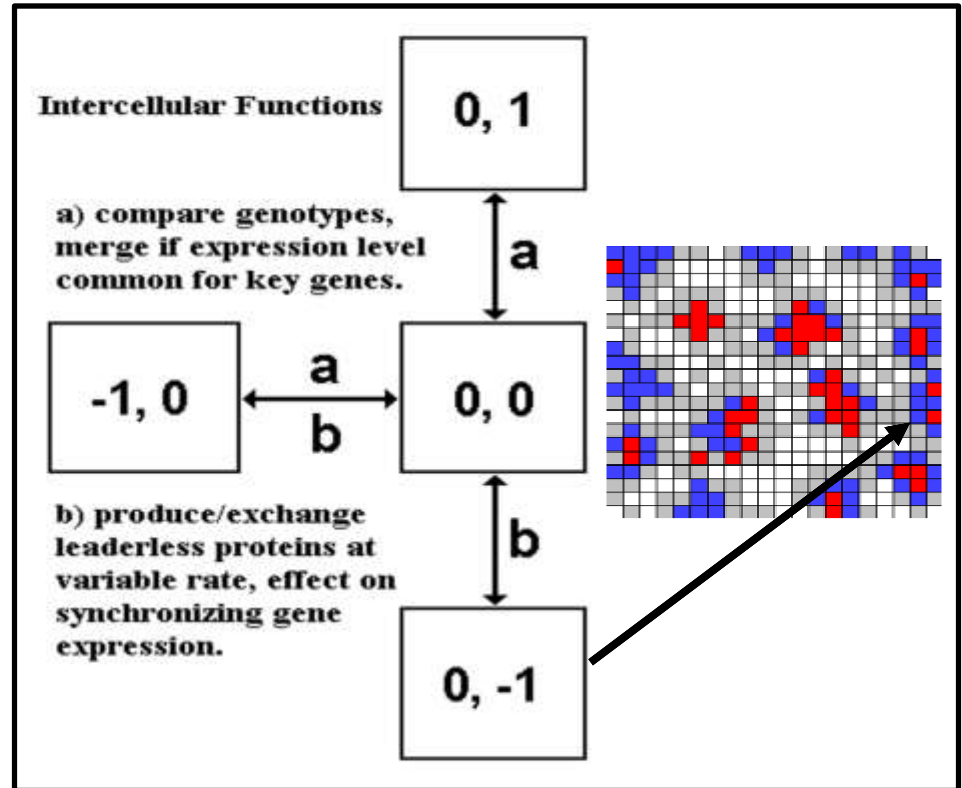
Evolving CAs: evolution of development (evo-devo) approach.

How to model intercellular signaling:

•secretion factors = leaderless proteins that influence neighboring cells, involved in a host of pathologies.

* colony formation is key component of self-renewal maintenance (pluripotent cells), full electrophysiologic maturity (neuronal cells).



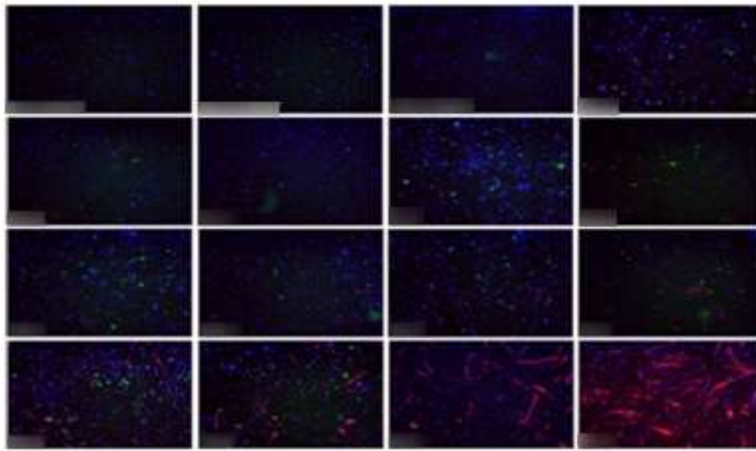


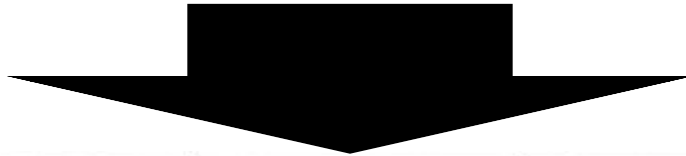
How do we understand transformation given infection in a cell population?

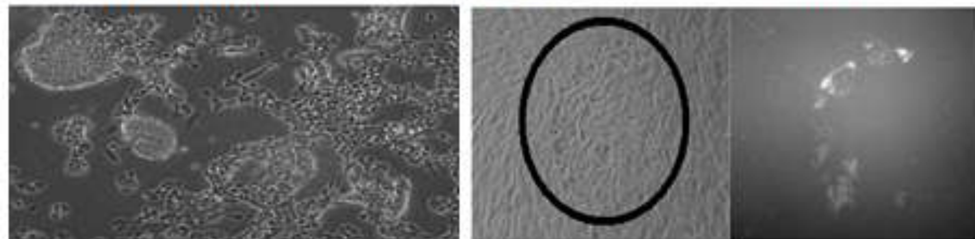
How do we understand the action of intercellular signaling (in this case secretory activity)?

* as a conditional probability? As a systems-level phenomenon?

* as nearest-neighbor influence (as opposed to mass action kinetics)?







Diverse Inputs:

* fibroblasts from different tissue types, species.

•individuality – some may be better than others, worse than others.

* different ambient conditions (niche, growth conditions, etc).

Diverse Outputs:

* "idealized" iPS cells.

* piPS = partially pluripotent cells.

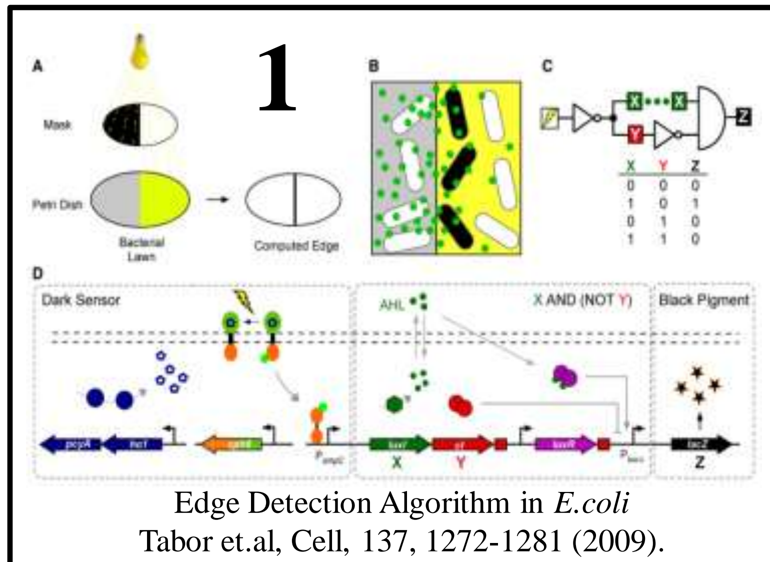
* piPS cells with surrounding differentiation (or carcinoma).

What are future directions for integrating models with biological data?

1) break biological processes down into computations (edge detection by cells).

2) use a time-course approach to find dynamic patterns during process, biological control modeling of RNA.

3) data integration – gene expression, methylation, proteomics – are there coherent patterns that can be exploited?



Edge Detection Algorithm in *E.coli*
Tabor et.al, Cell, 137, 1272-1281 (2009).