Cell reprogramming modelled as transitions in a hierarchy of cell cycles

Ryan Hannam, Alessia Annibale, Reimer Kühn

Disordered Systems Group
Department of Mathematics
King's College London

Biophysics of Perception

TU München — 23 October 2017

Outline

Outline

Cell Reprogramming

- Gurdon's nuclear transfer method

  Gurdon, Development (1962)

  - Transferred the nucleus of mature differentiated intestinal cell of a frog into an egg cell
  - Cell developed into a normal tadpole, then adult frog.
  - $\Rightarrow$ Genome of a differentiated cell contains all information needed to drive development into all different cells of an organism.

Cell Reprogramming

- Takahashi & Yamanaka cell reprogramming experiments

  Takahashi, Yamanaka, Cell (2006)

  - Introduce transcription factors (TFs) that are highly expressed in embryonic stem cells into fully differentiated cells
  - $\Rightarrow$ Transformed differentiated cells into pluripotent stem cells.
  - Managed to narrow down initial number of factors thought to be relevant (24) to just 4.
  - Now referred to as Yamanaka factors Oct3/4, Sox2, Klf4 and c-Myc

- Gurdon and Yamanaka shared the 2012 Nobel Prize for Medicine or Physiology

  "... for their discovery that mature cells can be reprogrammed to become pluripotent."

Cell Reprogramming

- Reprogramming (à la Takahashi-Yamanaka)
  - requires $\mathcal{O}(10)$ days for cells to be reprogrammed
  - ... with cells exhibiting intermediate 'chimera' profiles
  - is inefficient (just 0.001% success rate)
  - easier for cells higher up the differentiation cascade
  - now accepted to be achievable for all differentiated cell types
  - detailed mechanisms still poorly understood

Cell Reprogramming

- Reprogramming (à la Takahashi-Yamanaka)
  - requires $\mathcal{O}(10)$ days for cells to be reprogrammed
  - ... with cells exhibiting intermediate 'chimera' profiles
  - is inefficient (just 0.001% success rate)
  - easier for cells higher up the differentiation cascade
  - now accepted to be achievable for all differentiated cell types
  - detailed mechanisms still poorly understood
- Workshop Series at KCL (2009/10)

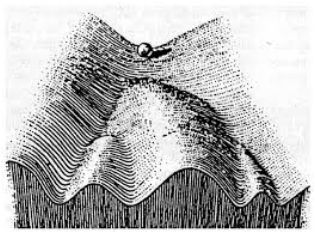
Cell Reprogramming

- Reprogramming (à la Takahashi-Yamanaka)
  - requires $\mathcal{O}(10)$ days for cells to be reprogrammed
  - ... with cells exhibiting intermediate 'chimera' profiles
  - is inefficient (just 0.001% success rate)
  - easier for cells higher up the differentiation cascade
  - now accepted to be achievable for all differentiated cell types
  - detailed mechanisms still poorly understood
- Workshop Series at KCL (2009/10)

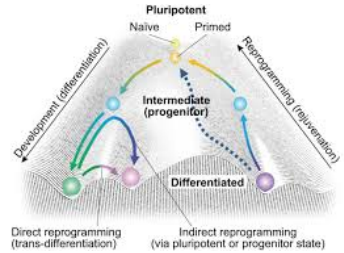
$\Rightarrow$ **... we should perhaps try to properly understand cell programming, before talking of cell re-programming.**

Cell Programming

- Waddington (1957)
  - Epigenetic Landscape



Waddington, *The Strategy of the Genes* (1957)



Takahashi & Yamanaka, Development (2015)

  - $\Leftrightarrow$ metaphorical explanation of cell differentiation and stability
- Kauffman (1969)

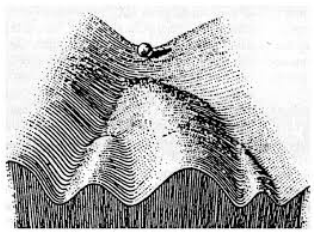
  J Theor. Biol. (1969)
  - Dynamics of gene expression levels as (random) Boolean network

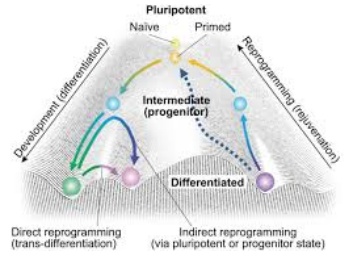
$$n_i(t+1) = f_i(\{n_j(t)\}_{j\in\partial i}) \ , \qquad n_i \in \{0,1\}$$

  - Rationalizes cell types as attractors of a dynamical system of many interacting degrees of freedom
  - Cell differentiation as transitions between attractors ??

Cell Programming

- Waddington (1957)
  - Epigenetic Landscape





Waddington, *The Strategy of the Genes* (1957) Takahashi & Yamanaka, Development (2015)

  - $\Leftrightarrow$ metaphorical explanation of cell differentiation and stability
- Kauffman (1969)

J Theor. Biol. (1969)

  - Dynamics of gene expression levels as (random) Boolean network

$$n_i(t+1) = f_i(\{n_j(t)\}_{j\in\partial i}) \ , \qquad n_i \in \{0,1\}$$

  - Rationalizes cell types as attractors of a dynamical system of many interacting degrees of freedom
  - Cell differentiation as transitions between attractors ??

Cell Programming

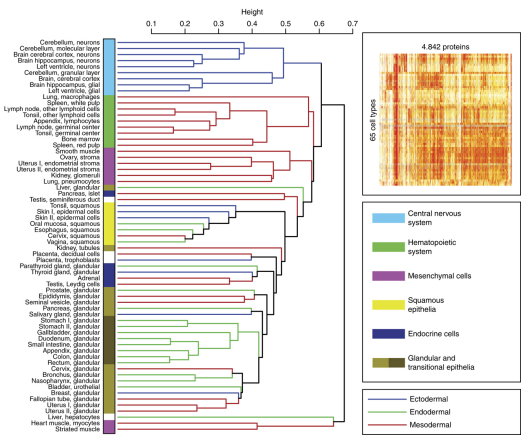
- Both Waddington and Kauffman address the problem of stability: different genetic make-ups give rise to same cell types
- Neither of them is very specific about mechanisms
  - Waddington: genes create the landscape (but how ?)
  - Kaufman: genes interact in complex ways (but why and how?)
- Both ideas very influential
  - Waddington: 3469 citations (GS)
  - Kauffman: 3651 citations (GS)
- Since 1980s: Analogy between attractor neural networks and gene regulatory networks
  - e.g. Derrida, Gardner & Zippelius (1987); Bastolla & Parisi (1998)
  - recently: reprogramming, Lang et al., PLoS Comp. Biol, (2014)

- Attractor models also largely ad hoc.

Stepping Back – The Phenomenology

- In humans: $\sim \mathbf{25,000}$ genes
  - $\mathcal{O}(60\%)$ typically expressed
    $\Rightarrow \mathcal{O}(10^{7,300})$ possible gene expression patterns
  - Yet only 300-400 cell types
  - Hierarchically organized

    – unavoidable consequence of evolutionary mechanisms (mutation, gene-exchange)
  - Cells are dynamical entities
    (cell cycles: $G_1 \to S \to G_2 \to M \to G_1 \ldots$)
    $\Leftrightarrow$ **at odds with energy landscapes**.

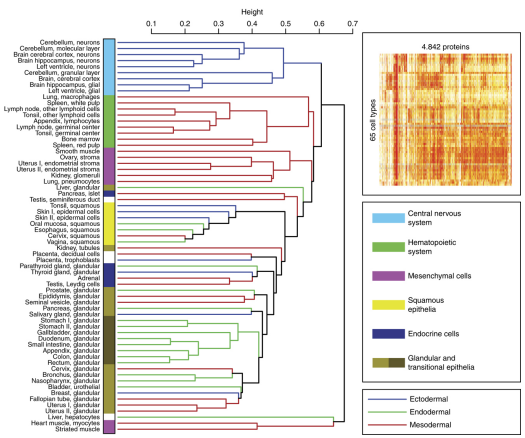


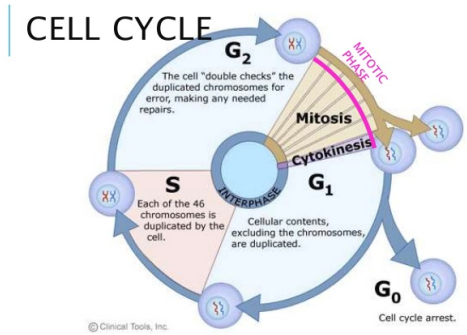
- **Q**: Does cell-chemistry provide mechanisms to generate this phenomenology?

Stepping Back – The Phenomenology

- In humans: $\sim 25,000$ genes
  - $\mathcal{O}(60\%)$ typically expressed
    $\Rightarrow \mathcal{O}(10^{7,300})$ possible gene expression patterns
  - Yet only 300-400 cell types
  - Hierarchically organized

    – unavoidable consequence of evolutionary mechanisms (mutation, gene-exchange)
  - Cells are dynamical entities
    (cell cycles: $G_1 \to S \to G_2 \to M \to G_1 \dots$)
    $\Leftrightarrow$ at odds with energy landscapes.



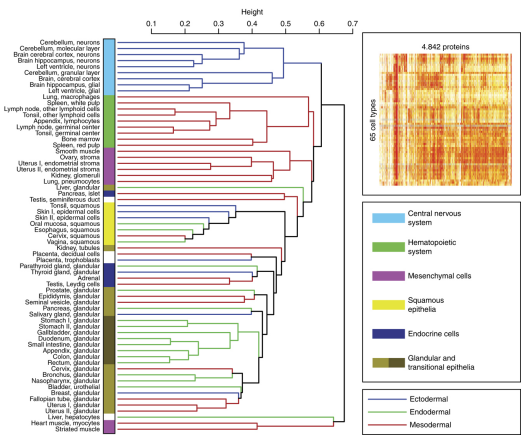


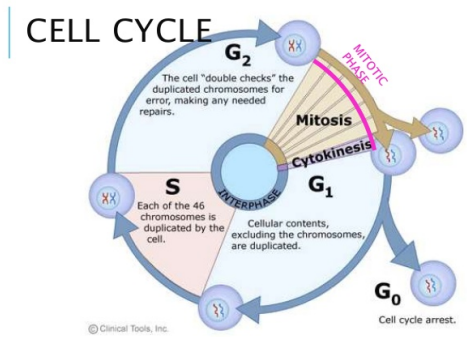
- Q: Does cell-chemistry provide mechanisms to generate this phenomenology?

Stepping Back – The Phenomenology

- In humans: $\sim \mathbf{25,000}$ genes
  - $\mathcal{O}(60\%)$ typically expressed
    $\Rightarrow \mathcal{O}(10^{7,300})$ possible gene expression patterns
  - Yet only 300-400 cell types
  - Hierarchically organized

    – unavoidable consequence of evolutionary mechanisms (mutation, gene-exchange)
  - Cells are dynamical entities
    (cell cycles: $G_1 \to S \to G_2 \to M \to G_1 \dots$)
    $\Leftrightarrow$ **at odds with energy landscapes**.



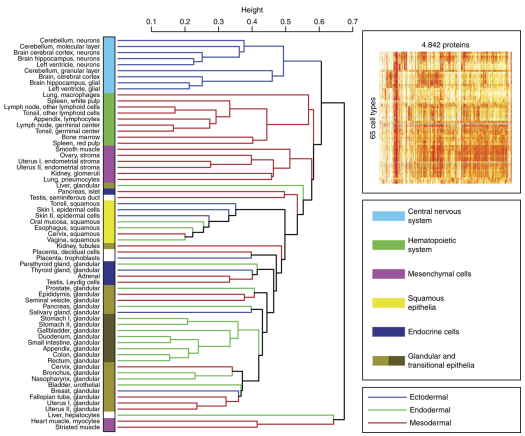


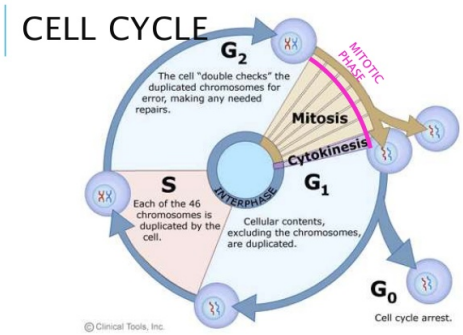
- **Q**: Does cell-chemistry provide mechanisms to generate this phenomenology?

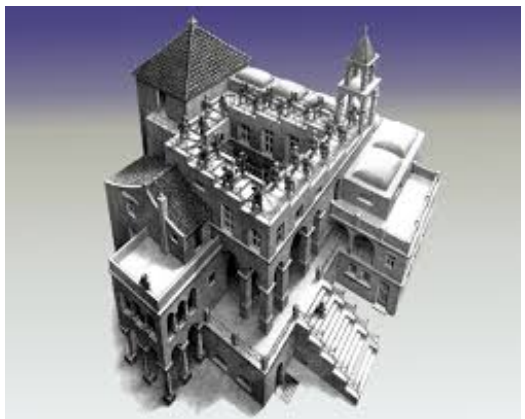
Stepping Back – The Phenomenology

- In humans: $\sim \mathbf{25,000}$ genes
  - $\mathcal{O}(60\%)$ typically expressed
    $\Rightarrow \mathcal{O}(10^{7,300})$ possible gene expression patterns
  - Yet only 300-400 cell types
  - Hierarchically organized

    – unavoidable consequence of evolutionary mechanisms (mutation, gene-exchange)
  - Cells are dynamical entities
    (cell cycles: $G_1 \to S \to G_2 \to M \to G_1 \dots$)
    $\Leftrightarrow$ **at odds with energy landscapes**.





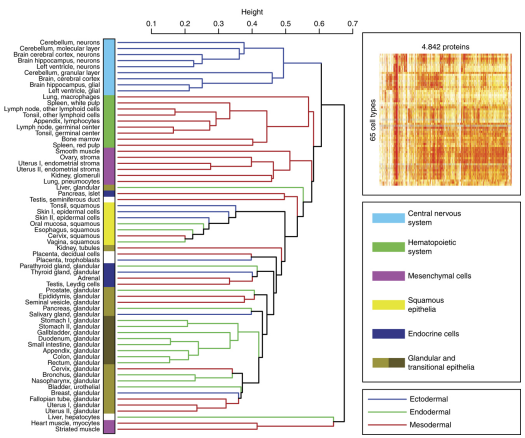


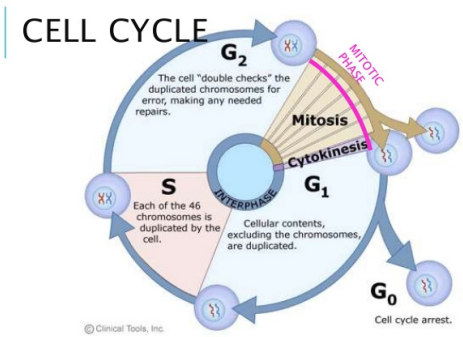
- **Q**: Does cell-chemistry provide mechanisms to generate this phenomenology?

Stepping Back – The Phenomenology

- In humans: $\sim 25,000$ genes
  - $\mathcal{O}(60\%)$ typically expressed
    $\Rightarrow \mathcal{O}(10^{7,300})$ possible gene expression patterns
  - Yet only 300-400 cell types
  - Hierarchically organized

    – unavoidable consequence of evolutionary mechanisms (mutation, gene-exchange)
  - Cells are dynamical entities
    (cell cycles: $G_1 \to S \to G_2 \to M \to G_1 \dots$)
    $\Leftrightarrow$ **at odds with energy landscapes**.





- **Q**: Does cell-chemistry provide mechanisms to generate this phenomenology?

Stepping Back – A Gedanken-Experiment

- Suppose I knew everything about cell chemistry, and I mean **really everything**!
- I would write down the complete set of equations describing all chemical reactions and physical processes occurring in a cell.

  (genes, mRNA, tRNA, μ-RNA, proteins, metabolites ... include effects of diffusion, small numbers ...).
- Suppose that I would integrate out all degrees of freedom except gene expression levels from my equations.
- Which properties would the reduced model necessarily have?
- It would
  - exhibit interactions between gene expression levels,
  - exhibit a non-Markovian dynamics.

- **BOTH properties necessary** in order to create cell-cycle type attractors with varying durations for the cycle states.

Stepping Back – A Gedanken-Experiment

- Suppose I knew everything about cell chemistry, and I mean **really everything**!
- I would write down the complete set of equations describing all chemical reactions and physical processes occurring in a cell,

  (genes, mRNA, tRNA, μ-RNA, proteins, metabolites ... include effects of diffusion, small numbers ...).
- Suppose that I would integrate out all degrees of freedom except gene expression levels from my equations.
- Which properties would the reduced model necessarily have?
- It would
  - exhibit interactions between gene expression levels
  - exhibit a non-Markovian dynamics

- **BOTH properties necessary** in order to create cell-cycle type attractors with varying durations for the cycle states.

Outline

A Minimal Reduced Model

- Simplifications
  - binary gene-expression levels $n_i \in \{0, 1\}$, $i = 1, \ldots, N$
  - discrete time counting stages of cell cycle (G_1, S, G_2, ...)
  - $\Rightarrow$ allows to ignore memory in interactions
- Dynamics

$$n_i(t+1) = \Theta\left[h_i(t) - \vartheta_i - \xi_i(t)\right]$$

  - with a 'local field' $h_i(t)$ encoding effects of interactions

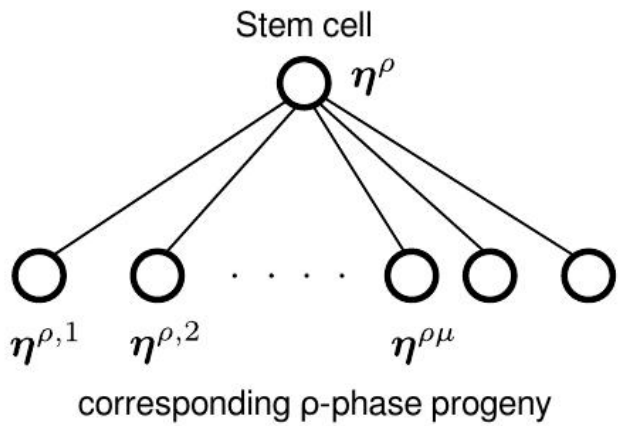
$$h_i(t) = \sum_j J_{ij} n_j(t) + \sum_{j,k} J_{ijk} n_j(t) n_k(t) + \ldots$$

    note: includes binary, ternary, quaternary ... interactions.
  - and $\xi_i(t)$ a noise-term, e.g. 'thermal'

$$\text{Prob}(\xi \leq x) = \frac{1}{2}\left(1 + \tanh\left(\frac{\beta}{2} x\right)\right)$$

A Minimal Reduced Model

- Further simplifications:
  - pair interactions only
  - uniform thresholds $\vartheta_i \equiv \vartheta$
- Construct interactions to
  - encode set of **hierarchically organized cell cycles**.
  - simplify to two-level hierarchy
    - $\boldsymbol{\eta}^{\rho}$, $\rho = 1, \ldots, C$ cycle state of pluripotent stem cell,
    - $\boldsymbol{\eta}^{\rho\mu}$, $\mu = 1, \ldots, M$ cycle state of μ-th daughter-cell



- stochastic synthetic setting: $\eta_i^{\rho}, \eta_i^{\rho\mu}$ random in $\{0, 1\}$
  relation $\boldsymbol{\eta}^{\rho} \to \boldsymbol{\eta}^{\rho\mu}$ probabilistic: $W_{\eta_i^{\rho} \to \eta_i^{\rho\mu}}$

A Minimal Reduced Model

- Take inspiration from neural networks
  - Storage of hierarchically organized patterns
    Parga, Virasoro, J. Phys.(Paris) (1986); Bös, RK, van Hemmen, Z. Phys. B (1988);
    Krogh, Herz, J.Phys. A (1988)
  - Storage of cyclic attractors
    Sompolinsky, Kanter, PRL (1986); Herz, Sulzer, RK, van Hemmen EPL (1988) & Biol. Cyb. (1989)

$$J_{ij} = \frac{1}{N}\sum_{\rho=1}^{C}\left\{\frac{(\eta_i^{\rho+1}-a^{\rho+1})(\eta_j^{\rho}-a^{\rho})}{a^{\rho}(1-a^{\rho})} + \sum_{\mu=1}^{M}\frac{(\eta_i^{\rho+1,\mu}-a_\mu(\eta_i^{\rho+1}))(\eta_j^{\rho\mu}-a_\mu(\eta_j^{\rho}))}{a^{\rho\mu}(1-a^{\rho\mu})}\right\}.$$

with

$$a^{\rho}=\mathbb{E}[\eta^{\rho}]\ ,\quad a^{\rho\mu}=\mathbb{E}[\eta^{\rho\mu}]\ ,\quad a_\mu(\eta^{\rho})=\mathbb{E}[\eta^{\rho\mu}|\eta^{\rho}]$$

- For details, see R Hannam, RK, and A Annibale, J. Phys. A **50** 425601 (2017)

Solving the Dynamics

- For $N \gg 1$ and $M/N \ll 1$, dynamics can be solved in closed form in terms of the order parameters,

$$\widetilde{m}_\rho(t) = \frac{1}{N}\sum_{i=1}^{N} \frac{\eta_i^\rho - a^\rho}{a^\rho(1-a^\rho)} n_i(t) \;, \quad \widetilde{m}_{\rho\mu}(t) = \frac{1}{N}\sum_{i=1}^{N} \frac{\eta_i^{\rho\mu} - a_\mu(\eta_i^\rho)}{a^{\rho\mu}(1-a^{\rho\mu})} n_i(t) \,,$$

similarity measures between system state $\boldsymbol{n}(t) = (n_i(t))$ and gene expression profiles of stem-cell/progenitor cycle states $\boldsymbol{\eta}^\rho$ and $\boldsymbol{\eta}^{\rho\mu}$, respectively.

$$\begin{aligned}
\widetilde{m}_\rho(t+1) &= \frac{1}{2}\left\langle \frac{\eta^\rho - a^\rho}{a^\rho(1-a^\rho)} \tanh\left(\frac{\beta}{2}\Big(h(t)-\vartheta\Big)\right)\right\rangle_{\boldsymbol{\eta}^\rho,\boldsymbol{\eta}^{\rho\mu}} \,, \\
\widetilde{m}_{\rho\mu}(t+1) &= \frac{1}{2}\left\langle \frac{\eta^{\rho\mu} - a_\mu(\eta^\rho)}{a^{\rho\mu}(1-a^{\rho\mu})} \tanh\left(\frac{\beta}{2}\Big(h(t)-\vartheta\Big)\right)\right\rangle_{\boldsymbol{\eta}^\rho,\boldsymbol{\eta}^{\rho\mu}}
\end{aligned}$$

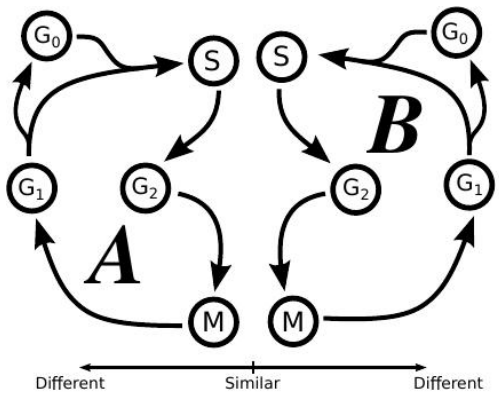
where

$$h(t) = \sum_\rho \left\{ (\eta^{\rho+1} - a^{\rho+1})\widetilde{m}_\rho(t) + \sum_\mu \left[\eta^{\rho+1,\mu} - a_\mu(\eta^{\rho+1})\right] \widetilde{m}_{\rho\mu}(t) \right\} .$$

Final Ingredient — A Conjecture

- We assume that there is (at least) one state of the cell cycle in which different cells have a relatively high degree of mutual similarity.
- Candidates are S (where DNA replication happens) and M (division of chromosomes, cytoplasm and organelles).

because both S and M deploy massive and sophisticated machineries unlikely to have been re-invented in different ways for different cell types of an organism.

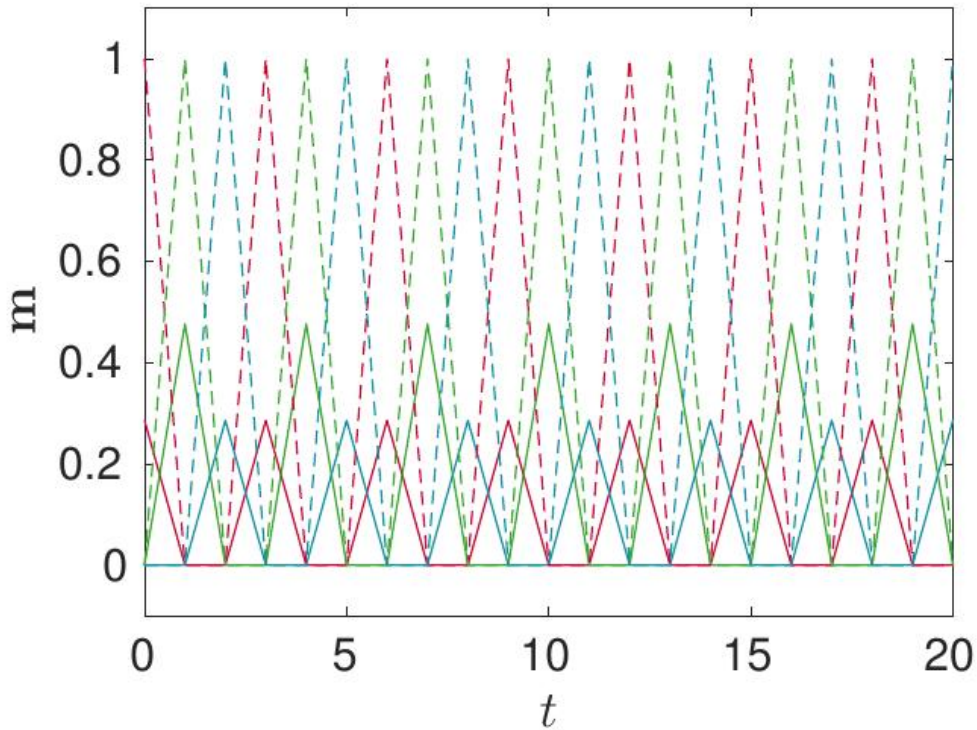


- States of high mutual similarity are natural targets for reprogramming protocols.
- A technical simplification: three state cycle, with one cycle state of greater mutual similarity than others.

Outline

Results

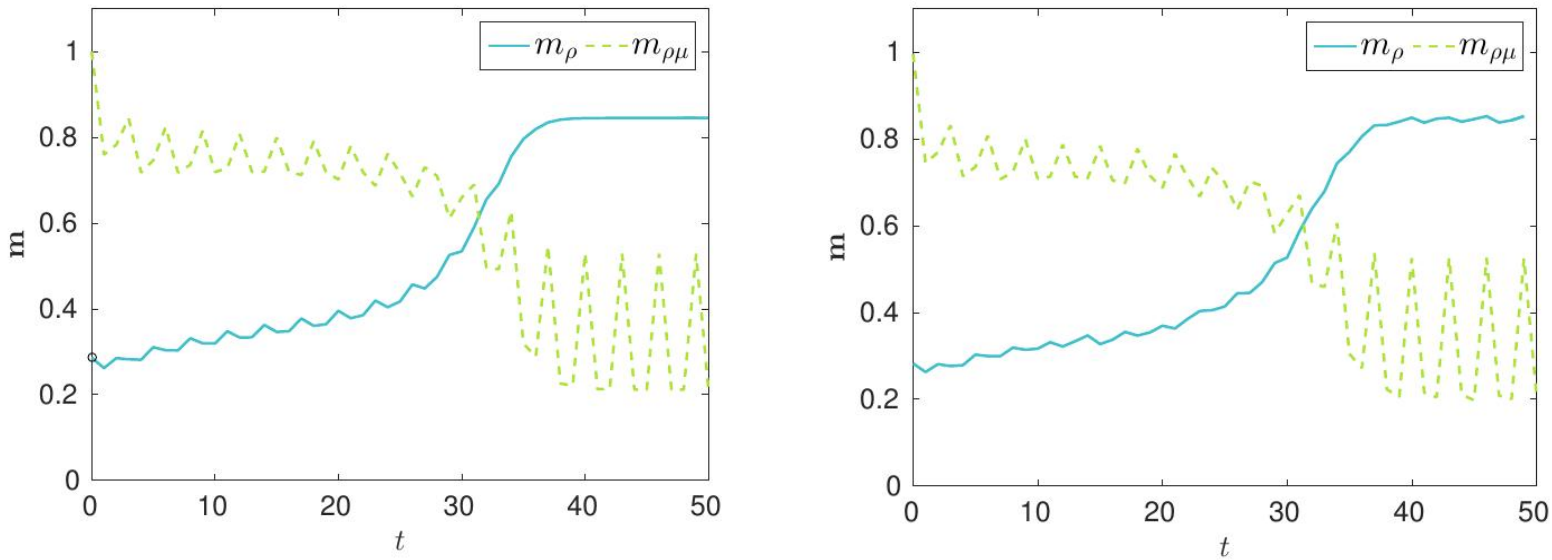
- Stable daughter cell cycle at low noise levels ($T = 0.01$)



Time dependent correlations with daughter cell cycle states (dashed) and stem-cell cycle states (full line); one of the three stem cell states has a high degree of similarity with the corresponding daughter cell state. R Hannam, RK, A Annibale, J Phys A (2017).

Results

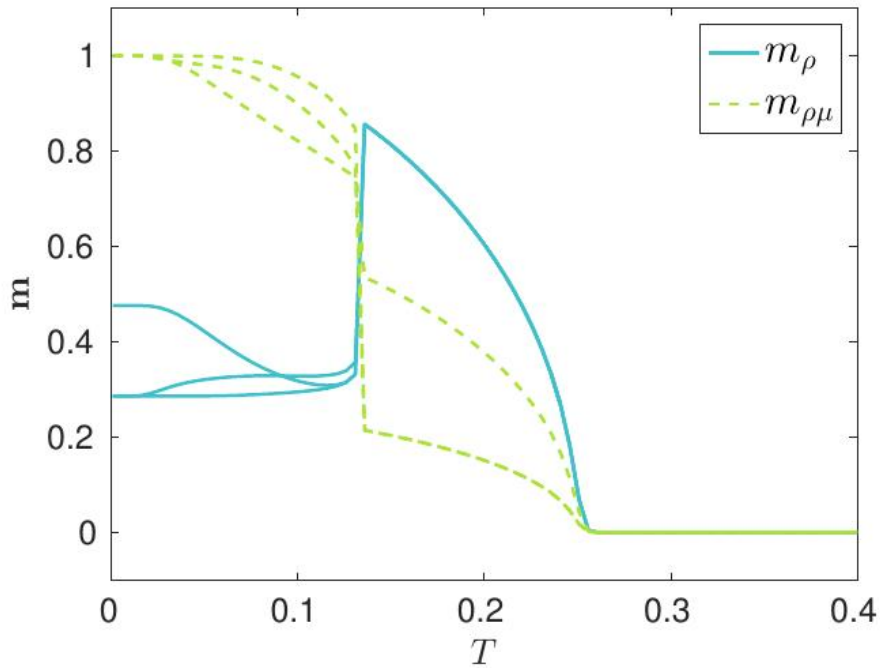
- Noise induced de-differentiation ($T = 0.14$)



Noise induced de-differentiation. The system is initialized in a daughter cell cycle. Only envelopes of correlations with cycle states are shown $m(t) = m_{\rho(t)}$, and $m(t) = m_{\rho(t),\mu}$. (Left: analytic results. Right: Simulation of a system of $N = 25,000$ genes). Note: de-differentiation takes $\mathcal{O}(10)$ cycles. R Hannam, RK, A Annibale, J Phys A (2017)

Results

- Noise-dependent stationary overlaps



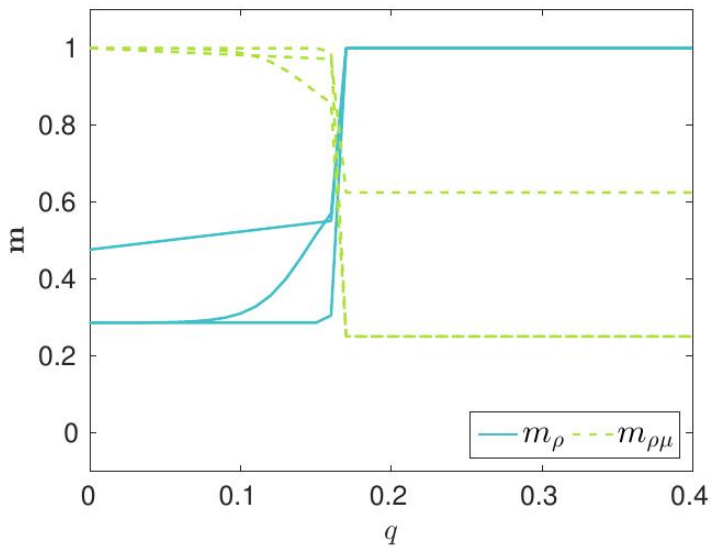
Asymptotic overlaps of the three stem-cell and daughter cell states as functions of the noise level. R Hannam, RK, A Annibale, J Phys A (2017).

Results

- Direct perturbation:
  - apply to progeny in cell state $(\bar{\rho}, \mu)$
  - to drive transition into cell state $\bar{\rho} + 1$

$$h_i(t) \to h_i(t) + c_i \; k \; (\eta_i^{\bar{\rho}+1} - a^{\bar{\rho}+1}) \widetilde{m}_{\bar{\rho}\mu}(t)$$

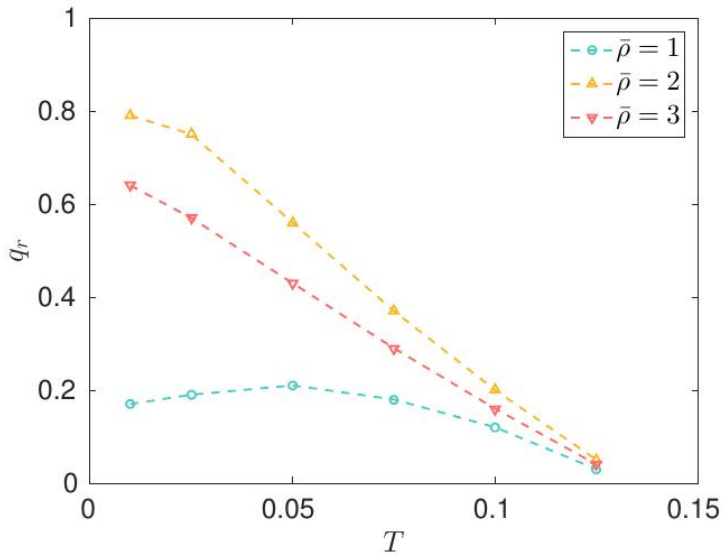
with $c_i = 1$ with probability q

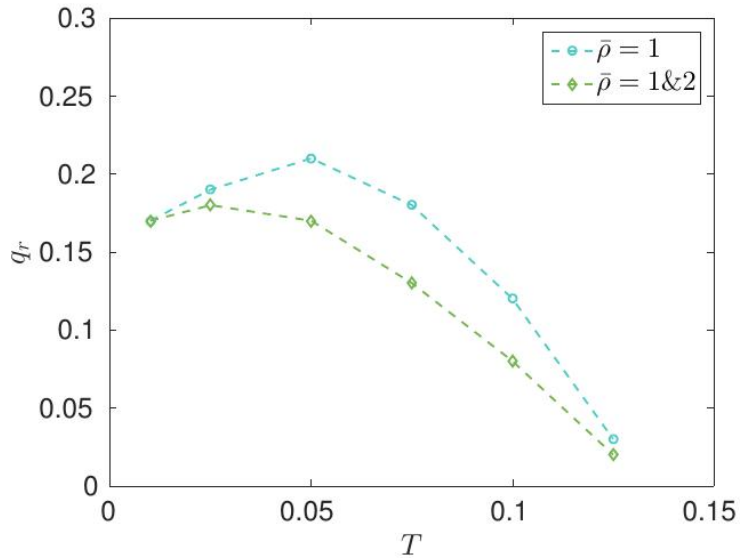


Asymptotic overlaps of the three stem-cell and daughter cell states as functions of the perturbation probability q at $T = 0.01$, $k = 1$. Perturbation applied prior to the most similar phase. R Hannam, RK, A Annibale, J Phys A (2017).

Results

- Temperature dependence of critical reprogramming fraction q_r.





Noise dependence of critical perturbation probability q_r at, $k = 1$. Left: perturbation applied to one phase. Right: Perturbation applied to two successive phases. R Hannam, RK, A Annibale, J Phys A (2017).

Results

- Critical fraction of genes to perturb: $q_r N \simeq$ 2,500 - 5,000 genes.
  ($\gg 4$ Yamanaka factors) **Disappointment!?**
- **Think again!** $\Leftrightarrow$ **Some Numerology:**
  - We should really be thinking of N as number of regulatory genes.
    $N \simeq 0.1 N_{\text{genes}} \simeq 2,500 \quad \Rightarrow \quad q_r N \simeq$ 250 - 500 genes
  - regulatory genes code for transcription factors (TFs)
  - each gene regulated by $\mathcal{O}(10)$ TFs
  - each TF involved in regulating $\mathcal{O}(100)$ genes
  - $\Rightarrow$ can perturb $q_r N$ genes with $\mathcal{O}(3-5)$ TFs! (Yamanaka territory!)
- A final note of caution: before rejoicing, note that with 2,500 genes, higher order interactions may be required to produce 300-400 stable cell types (and one can argue they must exist). This may change numbers again. Due to the collective nature of processes, changing details need not drastically change phenomenology and orders of magnitude, though. Inference on real data!

Results

- Critical fraction of genes to perturb: $q_r N \simeq$ 2,500 - 5,000 genes.
  ($\gg 4$ Yamanaka factors) **Disappointment!?**

- **Think again!** $\Leftrightarrow$ **Some Numerology:**
  - We should really be thinking of N as number of regulatory genes.
    $N \simeq 0.1 N_{\text{genes}} \simeq 2,500 \quad \Rightarrow \quad q_r N \simeq$ 250 - 500 genes
  - regulatory genes code for transcription factors (TFs)
  - each gene regulated by $\mathcal{O}(10)$ TFs
  - each TF involved in regulating $\mathcal{O}(100)$ genes

  - $\Rightarrow$ can perturb $q_r N$ genes with $\mathcal{O}(\mathbf{3-5})$ TFs! (Yamanaka territory!)

- A final note of caution: before rejoicing, note that with 2,500 genes, higher order interactions may be required to produce 300-400 stable cell types (and one can argue they must exist). This may change numbers again. Due to the collective nature of processes, changing details need not drastically change phenomenology and orders of magnitude, though. Inference on real data!

Results

- Critical fraction of genes to perturb: $q_r N \simeq$ 2,500 - 5,000 genes.
  ($\gg 4$ Yamanaka factors) **Disappointment!?**

- **Think again!** $\Leftrightarrow$ **Some Numerology:**
  - We should really be thinking of N as number of regulatory genes.
    $N \simeq 0.1 N_{\rm genes} \simeq 2,500 \quad \Rightarrow \quad q_r N \simeq$ 250 - 500 genes
  - regulatory genes code for transcription factors (TFs)
  - each gene regulated by $\mathcal{O}(10)$ TFs
  - each TF involved in regulating $\mathcal{O}(100)$ genes

  - $\Rightarrow$ can perturb $q_r N$ genes with $\mathcal{O}(\mathbf{3-5})$ TFs! (Yamanaka territory!)

- A final note of caution: before rejoicing, note that with 2,500 genes, higher order interactions may be required to produce 300-400 stable cell types (and one can argue they must exist). This may change numbers again. Due to the collective nature of processes, changing details need not drastically change phenomenology and orders of magnitude, though. Inference on real data!

Results

- Critical fraction of genes to perturb: $q_r N \simeq$ 2,500 - 5,000 genes.
  ($\gg 4$ Yamanaka factors) **Disappointment!?**

- **Think again!** $\Leftrightarrow$ **Some Numerology:**
  - We should really be thinking of N as number of regulatory genes.
    $N \simeq 0.1 N_{\rm genes} \simeq 2,500 \quad \Rightarrow \quad q_r N \simeq$ 250 - 500 genes
  - regulatory genes code for transcription factors (TFs)
  - each gene regulated by $\mathcal{O}(10)$ TFs
  - each TF involved in regulating $\mathcal{O}(100)$ genes

  - $\Rightarrow$ can perturb $q_r N$ genes with $\mathcal{O}(\mathbf{3-5})$ TFs! (Yamanaka territory!)

- A final note of caution: before rejoicing, note that with 2,500 genes, higher order interactions may be required to produce 300-400 stable cell types (and one can argue they must exist). This may change numbers again. Due to the collective nature of processes, changing details need not drastically change phenomenology and orders of magnitude, though. Inference on real data!

Results

- Critical fraction of genes to perturb: $q_r N \simeq$ 2,500 - 5,000 genes.
  ($\gg 4$ Yamanaka factors) **Disappointment!?**

- **Think again!** $\Leftrightarrow$ **Some Numerology:**
  - We should really be thinking of N as number of regulatory genes.
    $N \simeq 0.1 N_{\rm genes} \simeq 2,500 \quad \Rightarrow \quad q_r N \simeq$ 250 - 500 genes
  - regulatory genes code for transcription factors (TFs)
  - each gene regulated by $\mathcal{O}(10)$ TFs
  - each TF involved in regulating $\mathcal{O}(100)$ genes

  - $\Rightarrow$ can perturb $q_r N$ genes with $\mathcal{O}(\mathbf{3-5})$ TFs! (Yamanaka territory!)

- A final note of caution: before rejoicing, note that with 2,500 genes, higher order interactions may be required to produce 300-400 stable cell types (and one can argue they must exist). This may change numbers again. Due to the collective nature of processes, changing details need not drastically change phenomenology and orders of magnitude, though. Inference on real data!

Outline

Summary

- Argued
  - that cell-chemistry gives rise to interactions and dynamics of the form needed to create dynamical attractors (stable cell-cycles),
  - that evolutionary mechanisms entail that cell-types are hierarchically organized.
- Constructed a minimal model with hierarchically organized cell cycles.
- Conjectured that (at least) one cycle state is similar in all cell types.
- Argued that this state is natural target for reprogramming protocols.
- Investigated reprogramming using a model inspired by NNs.
- Found that reprogramming takes several cycles.
- Found that reprogramming requires perturbations that can be achieved with realistic numbers of TFs.
- Next step: **inference on real data**; V. Pachnis @ Crick Institute