Biological Computation

Sara-Jane Dunn
Microsoft Research
@EssJayD
What about the 21st Century?
Starting MS-DOS...
C:\>_
Biological Computation

noun

The biochemical information-processing carried out by cells in order to transform chemical, electrical and mechanical cues into biological function.
Biological Computation at Microsoft Research

Domain Specific Languages

Programming DNA
Programming Genetic Devices
Reprogramming Cells

Abstractions, Compilers, Verification

DNA Domains
DNA Sequences
DNA Biophysics

Genetic Circuit
Chemical Reaction Network
Ordinary Differential Equations

Symbolic Regulatory Network
Labelled Transition System
Continuous Time Markov Chain

Cell Biophysics
Partial Differential Equations

Biological experiments as programs
Lab robotics

Automation, Analytics & Infrastructure
Embryonic stem cells (ESCs) are pluripotent, meaning they can generate all cell types of the adult body.
ESCs arise only transiently

We must understand how to direct differentiation
Induced Pluripotent Stem Cell (iPSC)

“Reprogramming”

Takahashi & Yamanaka, Cell (2006)
It’s not even that difficult...

...but it is inefficient, and poorly understood.
Stem cell utopia

Somatic cell → Induced Pluripotent Stem Cell (iPSC) → Various differentiated cells
We need to understand the biological program.
Modelling Dynamic Gene Interactions

Boolean Network

Kauffman, 1969

State transition system

Kauffman, 1969
Challenge: In reality, experimental data only indicates the possibility of an interaction. Classifying interactions as possible defines $2^n$ unique topologies.
Logical update functions

- Cannot define update functions dependent on named regulators
- Can consider whether some, all or none of the activators / repressors are active
- 18 possible rules cover all cases
- An ABN with $C$ components and $n$ possible interactions defines concrete Boolean Networks

\[ 2^n \times 18^C \]

Abstract

Boolean Network

At least one activator is on

All activators are on

All activators and at least one repressor are on
Challenge: Unique networks can produce the same behaviour

We need to consider the set of potential models
Examining Dynamic Behaviour

Concentration vs. Time

Simulation

State Transition System
From software verification to reprogramming cells.

1525 Started Multi Adder Test.

1545 Relay #70 Panel F
(moth) in relay.

Relay #70 Panel F

First actual case of bug being found.

95^/ 300.00
1000 Andam started.
13° C (O33) MP - MC
(033) PRO 2 2.130476415

1.2700 9.037847025
9.0378479545 comma
4.615925059 (2)
“If gene A is active, then gene B is active or gene C is active.”

“A or B is active, and C or D is active, and E is active, but A, D and E cannot all be active at the same time.”

“A or B is active, and A determines C, but B and C are both inactive.”

Borrowing Techniques From Software Verification

SAT problem

Z3
Reasoning Engine for Interaction Networks (RE:IN)

Abstract Boolean Network

Experimental observations

Concrete networks consistent with experimental observations

Yordanov & Dunn et al., npj Systems Biology and Applications (2016)

research.microsoft.com/rein
http://rein.cloudapp.net/
The biological program governing ‘reprogramming’

- Molecular program consistent with 149 experiments; predictive accuracy of 80%
- Increased reprogramming efficiency 400x, and reduced time to just 24hrs
- Predicted gene activation trajectories substantiated even at single cell resolution
- Elucidated that individual cells follow a deterministic trajectory
Acknowledgements

- Amy Li, Graziano Martello and Austin Smith
- Boyan Yordanov, Christoph Wintersteiger
- www.research.microsoft.com/rein
Biological Computation Group

Collaborators

Reprogramming Stem Cells
University of Cambridge: Austin Smith, Amy Li, Brian Hendrich, Nicola Reynolds, Bertille Montibus
University of Padova: Graziano Martello
Microsoft Research: Christoph Wintersteiger
Kings College London: Angela Oliveira Pisco, Fiona Watt
University of Toronto: Peter Zandstra

Programming DNA
University of Washington: Georg Seelig, Gourab Chatterjee, Suzie Pun
University of New Mexico: Matthew Lakin
Rice University: Dave Zhang
University of Cambridge: Ulrich Keyser, Elisa Hemmig
Microsoft Research: Karin Strauss, Yuan Chen
Caltech: Frits Dannenberg

Programming Genetic Devices
University of Cambridge: Jim Ajioka, Jim Haseloff, Om Patange, Eugene Nadezhdin
UCL: Chris Barnes, Luca Rosa

Core Platform
Microsoft Research: Helene Steiner, Agile Projects Team
Industry: Twist, Synthace

Part-time

Filippo Polo
Niall Murphy
James Locke
Reprogramming

“cocktail

Reference

Oct4, Sox2, Klf4, LIF

Takahashi & Yamanaka (2006)

Oct4, Sox2, Klf4, 2i+LIF

Silva et al. (2008)

Oct4, Sox2, Klf4, Esrrb, LIF

Sone et al. (2017)

Oct4, Esrrb, Nanog, Klf4

Feng et al. (2009)

Sox2, Sall4, Nanog, Klf4

Buganim et al. (2012)
INFORMATION-
PROCESSING

**Gen A** → **Gen C** ↓ **Gen B**

**INPUT**

**OUTPUT**

*Biological program*