Modelling and Analysis in Bioinformatics: Simulating dynamical network models

Antti Honkela

4 October 2016

Local overview of the course

- Past two weeks: abstract networks
- Upcoming two weeks: simulating dynamical network models and network inference
  - This week: How to simulate models of interacting entities?
  - Next week: How to infer unknown networks from observations?

This ain't no physics

- Biology does not follow universal laws.

> *"Nothing in Biology Makes Sense Except in the Light of Evolution"*
>
> —Theodosius Dobzhansky

This ain't no physics

- Biology does not follow universal laws.

  > *"Nothing in Biology Makes Sense Except in the Light of Evolution"*
  >
  > —Theodosius Dobzhansky

- We study biology through models, which are simplifications of the world.

  > *"Essentially, all models are wrong, but some are useful."*
  >
  > —George E. P. Box

Gene regulation?

What is gene regulation?

Why are genes regulated?

How are genes regulated?

How can gene regulation be modelled?

Gene regulation?

What is gene regulation?

Why are genes regulated?

How are genes regulated?

How can gene regulation be modelled?

Learning goals for this week

- To understand the *reasons for the complexity* of gene regulation.
- To know *modelling formalisms* used to model gene regulation.
- To be able to *simulate* gene expression and regulation.
- To understand the effects of system parameters through simulation.

Outline

Outline

Gene expression

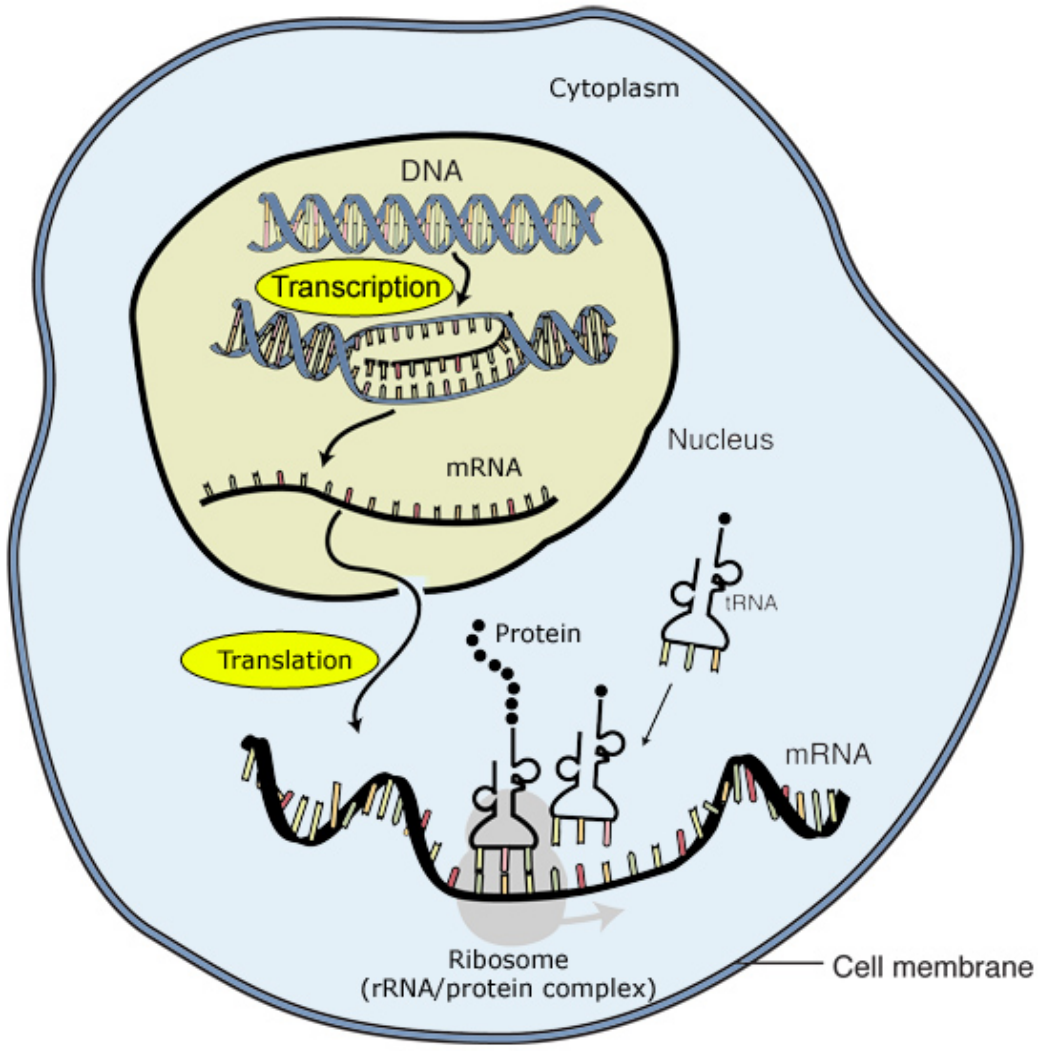
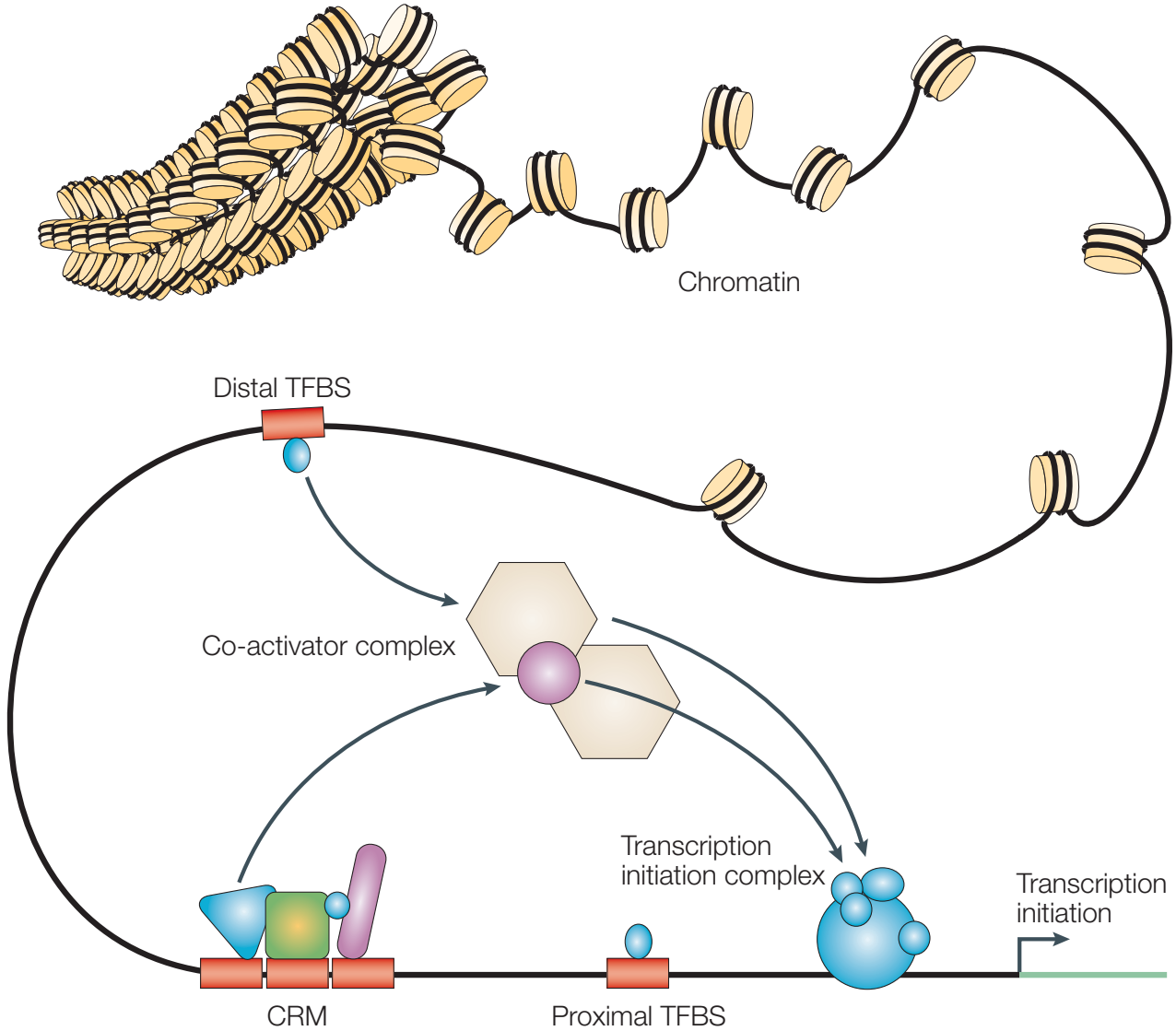


Image adapted from: National Human Genome Research Institute.

Transcrption regulation



(Image from: Wasserman & Sandelin. *Nat Rev Genet.* 5(4):276-87, 2004.)

Facts about gene regulation

- ≈20,000 genes in human, estimated 2000 of those are transcription factors (TFs)
- 250,000+ proteins
- Gene expression is highly spatially and temporally regulated
- E.g. nearly half of mouse genes vary by circadian rhythm somewhere in the body

The process and regulation of gene expression I

1. Recruitment and formation of transcription initiation complex
   - DNA accessibility, TF binding, epigenetics

The process and regulation of gene expression I

1. Recruitment and formation of transcription initiation complex
   - DNA accessibility, TF binding, epigenetics
2. Transcription initiation
   - TF binding, TF post-translational modifications

The process and regulation of gene expression I

1. Recruitment and formation of transcription initiation complex
   - DNA accessibility, TF binding, epigenetics
2. Transcription initiation
   - TF binding, TF post-translational modifications
3. Elongation
   - Pausing, polymerase falling off

The process and regulation of gene expression I

1. Recruitment and formation of transcription initiation complex
   - DNA accessibility, TF binding, epigenetics
2. Transcription initiation
   - TF binding, TF post-translational modifications
3. Elongation
   - Pausing, polymerase falling off
4. Transcription termination
   - Termination factors

The process and regulation of gene expression II

5. Splicing and other RNA processing
    - Splicing factors

The process and regulation of gene expression II

5. Splicing and other RNA processing
   - Splicing factors
6. mRNA transport
   - Various RNA-binding proteins

The process and regulation of gene expression II

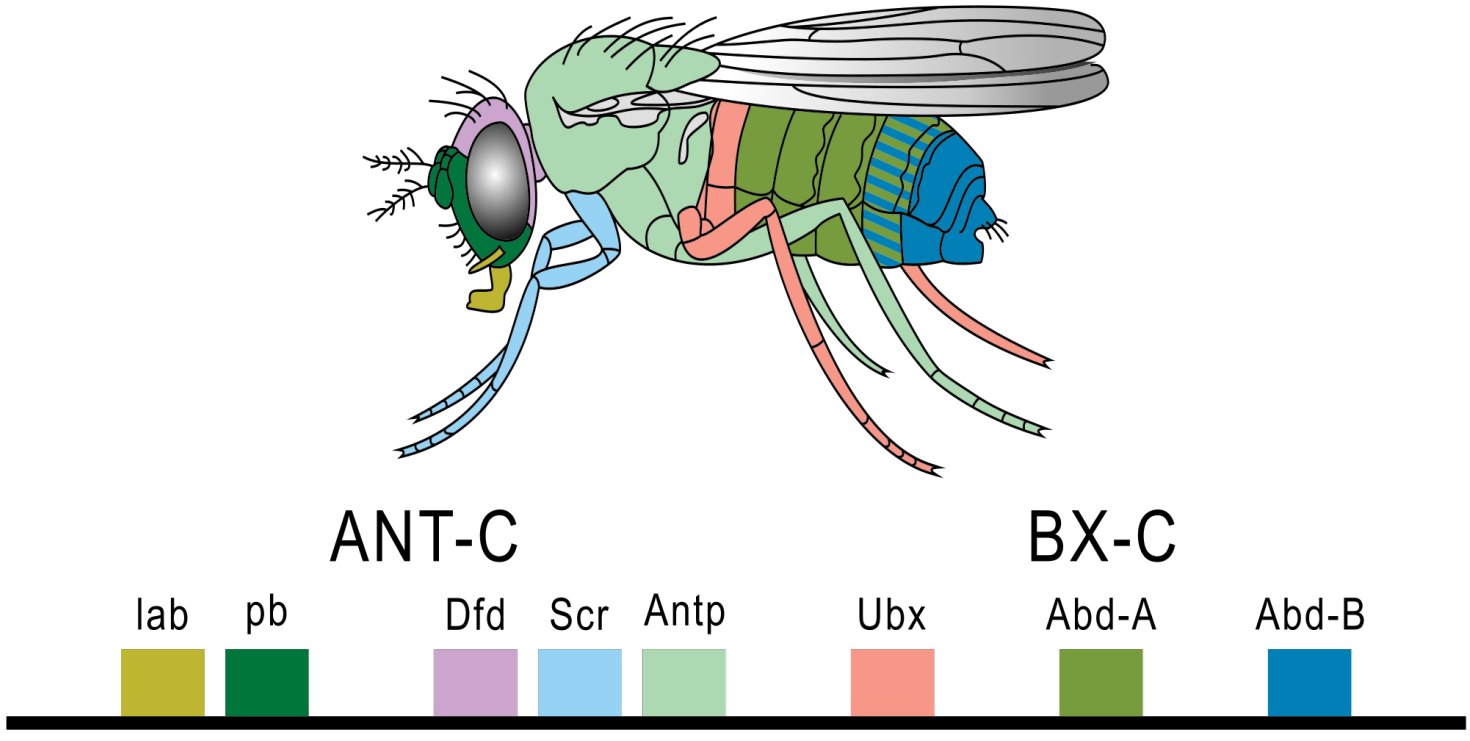
5. Splicing and other RNA processing
   - Splicing factors
6. mRNA transport
   - Various RNA-binding proteins
7. Translation
   - Ribosome function, tRNA availability

The process and regulation of gene expression II

5. Splicing and other RNA processing
   - Splicing factors
6. mRNA transport
   - Various RNA-binding proteins
7. Translation
   - Ribosome function, tRNA availability
8. mRNA degradation
   - Various mechanisms; incl. micro RNAs (miRNAs) and nonsense-mediated mRNA decay

Example: regulation of development

Example: regulation of development



Outline

A cell as a dynamical system

The state of the cell as a d-dimensional vector:

$$\mathbf{x}(t) = (x_1(t), \ldots, x_d(t))^T \in \mathcal{C} \subset \mathbb{R}^d$$

Questions:

- Which variables should the state include?
- What values are allowed for the variables?
- How to model the evolution of the state over time?
- What about spatial aspects? (Ignored here)

Different approaches for simulating a cell

Discrete state – Continuous state

Discrete time – Continuous time

Deterministic – Stochastic

Different approaches for simulating a cell

Discrete state – Continuous state

Discrete time – Continuous time

Deterministic – Stochastic

Different approaches for simulating a cell

Discrete state – Continuous state

Discrete time – Continuous time

Deterministic – Stochastic

Different approaches for simulating a cell

Discrete state – Continuous state

Discrete time – Continuous time

Deterministic – Stochastic

Different approaches for simulating a cell

Discrete state – Continuous state

Discrete time – Continuous time

Deterministic – Stochastic

Different approaches for simulating a cell

Discrete state – Continuous state

Discrete time – Continuous time

Deterministic – Stochastic

Different approaches for simulating a cell

Discrete state – Continuous state

Discrete time – Continuous time

Deterministic – Stochastic

Different approaches for simulating a cell

Discrete state – Continuous state

Discrete time – Continuous time

Deterministic – Stochastic

Different approaches for simulating a cell

Discrete state – Continuous state

Discrete time – Continuous time

Deterministic – Stochastic

An example system

Consider a system with two regulators (transcription factors) ANT and BEE regulating a target gene TAR:

$$\mathbf{x}(t) = \begin{pmatrix} x_{ANT}(t) \\ x_{BEE}(t) \\ x_{TAR}(t) \end{pmatrix}$$

Ordinary differential equations: Modelling a deterministic continuous time / continuous state process

A general form for a differential equation:

$$\frac{d\mathbf{x}(t)}{dt} = f(\mathbf{x}(t), t)$$

Ordinary differential equations: Modelling a deterministic continuous time / continuous state process

A general form for a differential equation:

$$\frac{d\mathbf{x}(t)}{dt} = f(\mathbf{x}(t), t)$$

Example of a regulatory model

$$f(\mathbf{x}(t), t) = f(\mathbf{x}(t)) = \begin{pmatrix} -\lambda_{ANT} x_{ANT}(t) \\ -\lambda_{BEE} x_{BEE}(t) \\ c \cdot x_{ANT}(t) x_{BEE}(t) - \lambda_{TAR} x_{TAR}(t) \end{pmatrix}$$

Simulating ordinary differential equations (ODEs)

Consider an *initial value problem*

$$\frac{d\mathbf{x}(t)}{dt} = f(\mathbf{x}(t), t), \quad \mathbf{x}(0) = \mathbf{x}_0$$

Numerical solution in theory (Euler's method):

- Pick a (small) time step δ
- Simulate a single discrete step:

$$\mathbf{x}(t + \delta) = \mathbf{x}(t) + \delta \cdot f(\mathbf{x}(t), t)$$

In practice:

- Use a ready ODE solver

Ordinary differential equation models in practice

Pros:

Cons:

Ordinary differential equation models in practice

Pros:

- Powerful framework
- Can define mechanistic models with interpretable parameters
- Efficient solvers widely available

Cons:

- Cannot model discrete variables and discrete transitions
- Deterministic model can be unrealistic

Modelling a stochastic continuous time / continuous state process

A general form for a stochastic differential equation:

$$d\mathbf{x}(t) = f(\mathbf{x}(t), t)dt + \Sigma(\mathbf{x}(t), t)dB(t)$$

where $dB(t)$ is a differential Brownian motion satisfying in a suitable sense (Itô integral)

$$\int_0^t dB(t) = N(0, t) \quad (cf. \int_0^t dt = t)$$

Simulating a stochastic continuous time / continuous state process

Consider an *initial value problem*

$$d\mathbf{x}(t) = f(\mathbf{x}(t), t)dt + \Sigma(\mathbf{x}(t), t)dB(t), \quad \mathbf{x}(0) = \mathbf{x}_0$$

Numerical solution in theory (Euler-Maruyama method):

- Pick a (small) time step δ
- Simulate a single discrete step:

$$\mathbf{x}(t + \delta) = \mathbf{x}(t) + \delta \cdot f(\mathbf{x}(t), t) + \Sigma(\mathbf{x}(t), t)\eta_t,$$

where

$$\eta_t \sim N(0, \delta\mathbf{I})$$

Stochastic differential equation models in practice

Pros:

Cons:

Stochastic differential equation models in practice

Pros:

- Powerful framework
- Can define mechanistic models with interpretable parameters
- Can provide very realistic stochastic models of continuous phenomena

Cons:

- Cannot model discrete variables and discrete transitions
- Solvers not as widely available as ODE solvers

Discrete states and Markov property

- Some properties are inherently discrete (e.g. number of molecules x in a single cell)
- Can have a big impact on modelling small non-negative values
- For discrete states and jumps, duration of a state becomes important
- Common simplification: Markov property

$$p(\mathbf{x}(t+\delta)|\mathbf{x}(\tau \leq t)) = p(\mathbf{x}(t+\delta)|\mathbf{x}(t))$$

Markov jump processes: Modelling a stochastic continuous time / discrete state process

$$p(\mathbf{x}(t+dt)|\mathbf{x}(t)) = f(\mathbf{x}(t)), \quad \mathbf{x}(0) = \mathbf{x}_0$$

Simulation loop:

- Simulate the time until next change (depends on f)
- Simulate the change itself

Try it out in practice in the exercises!

Markov jump processes in practice

Pros:

Cons:

Markov jump processes in practice

Pros:

- Powerful framework
- Can define mechanistic models with interpretable parameters
- Can provide very realistic stochastic models of discrete phenomena

Cons:

- General purpose solves usually not available (but not too difficult to implement)
- Simulation can be really slow if many potential states

How to select the function f?

- Selecting f is a big modelling choice along with selecting the approach
- Assuming a well-mixed system of discrete entities one can derive an f for Markov jump process
- The average of this process follows a known ODE
- Stochasticity can be captured better by a SDE
- More about these in the study circle and the exercises

Outline

Learning goals for this week

- To understand the *reasons for the complexity* of gene regulation.
- To know *modelling formalisms* used to model gene regulation.
- To be able to *simulate* gene expression and regulation.
- To understand the effects of system parameters through simulation.

Next steps

Thursday:

- Study circle on simulation methods
- Computer exercises on simulation methods

Next week:

- Network inference

Tasks for the study circle on Thursday

Papers:

- Gillespie D. Exact stochastic simulation of coupled chemical reactions. *Journal of physical chemistry* 81(25):2340–2361, 1977.
- Gillespie D. The chemical Langevin equation. *The journal of chemical physics* 113(1):297–306, 2000.

Tasks for different groups:

Tasks for the study circle on Thursday

Papers:

- Gillespie D. Exact stochastic simulation of coupled chemical reactions. *Journal of physical chemistry* 81(25):2340–2361, 1977.
- Gillespie D. The chemical Langevin equation. *The journal of chemical physics* 113(1):297–306, 2000.

Tasks for different groups:

Group 1 (Last name starts with A-H)	Read Gillespie (1977) (especially Secs. I, IIIB, IIIC)
Group 2 (Last name starts with I-L)	Read Gillespie (2000) (especially Secs. I, II, III)
Group 3 (Last name starts with M-Z)	Read Gillespie (2000) (especially Secs. I, II, IV)