Modelling and Analysis in Bioinformatics: Gene regulation and experimental functional genomics

Antti Honkela

9 November 2015

Local overview of the course

- Past two weeks: abstract networks
- Upcoming two weeks: concrete gene regulatory networks
  - This week: How does gene regulation manifest in gene expression?
  - 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?

Learning goals for this week

- To understand the *objectives* of gene regulation.
- To recognise *basic mechanisms* of gene regulation.
- To be able to *simulate* gene expression and regulation.
- To understand the effects of system parameters through simulation.

Outline

Outline

- Motivation
- Gene regulation overview
- Mechanics of gene expression
- Examples of regulatory mechanisms
- Experimental functional genomics
- Simulating gene regulation
- Conclusion and next steps

Gene expression

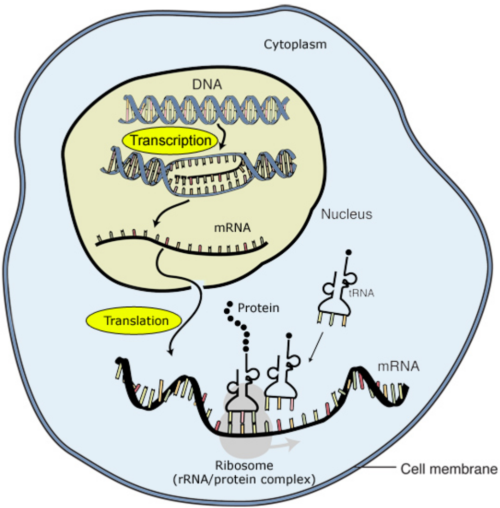
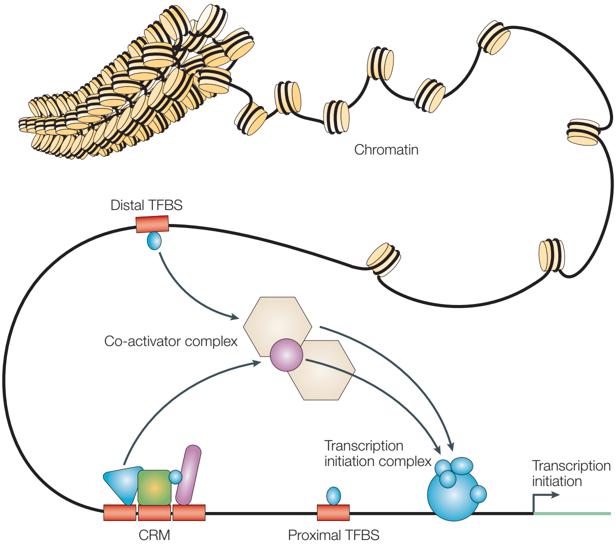


Image adapted from: National Human Genome Research Institute.

Transcrption regulation



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

Facts about gene regulation

- $\approx$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

E. coli

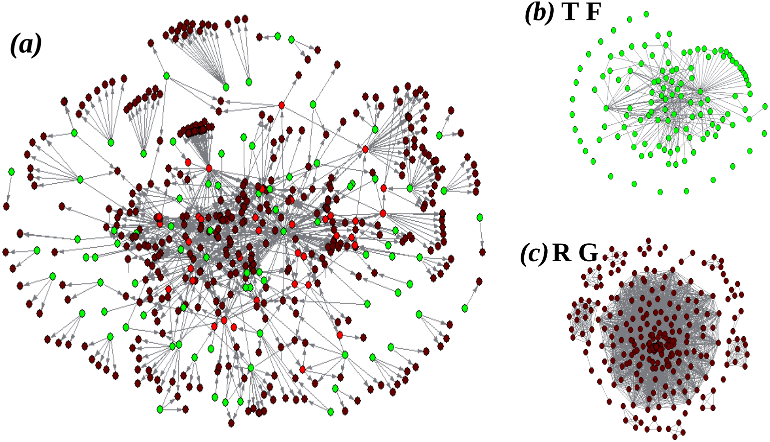
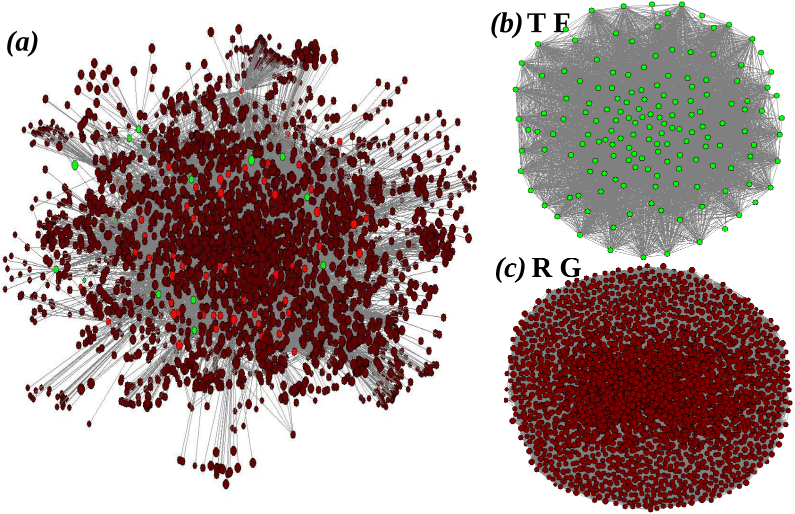


Figure from Guzmán-Vargas & Santillán (2008), doi:10.1186/1752-0509-2-13, CC-BY

S. cerevisiae



Outline

Steps in gene expression

1. Recruitment and formation of transcription initiation complex
2. Transcription initiation
3. Elongation
4. Transcription termination
5. Splicing and other RNA processing
6. mRNA transport
7. Translation
8. mRNA degradation

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

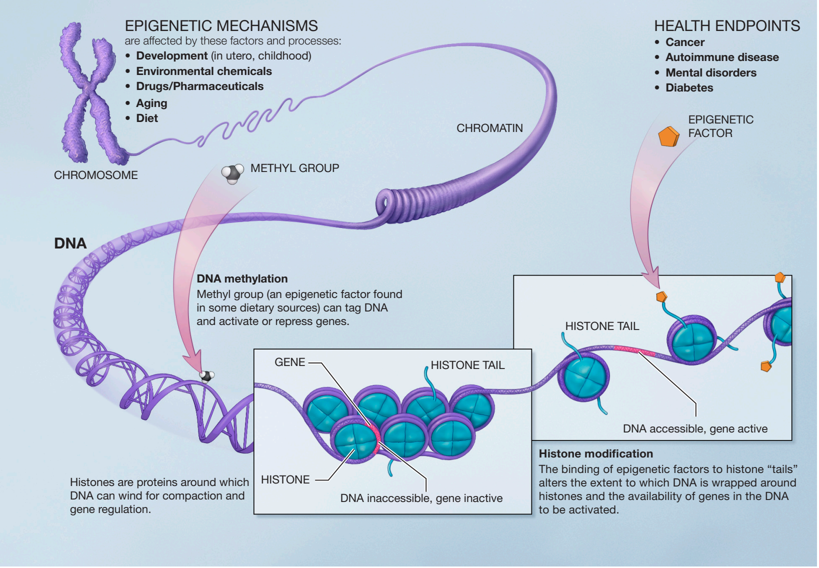
Regulation of gene expression II

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

Steps in gene expression

1. Recruitment and formation of transcription initiation complex
2. Transcription initiation
3. Elongation
4. Transcription termination
5. Splicing and other RNA processing
6. mRNA transport
7. Translation
8. mRNA degradation

Epigenetics



TF binding

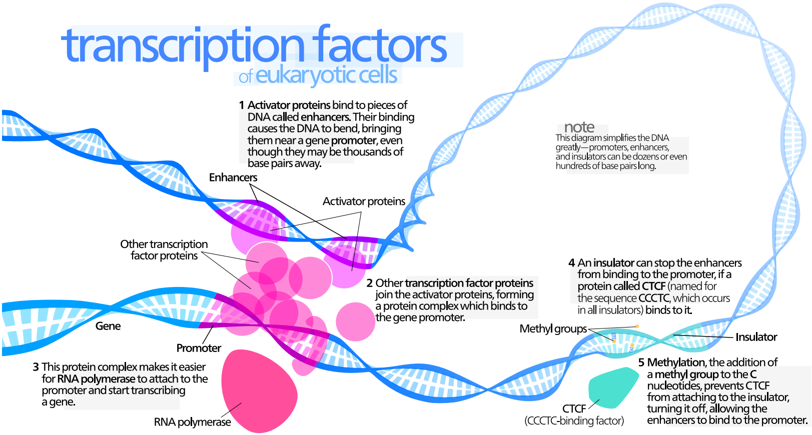


Figure by Kelvinsong, `https://commons.wikimedia.org/wiki/File:Transcription_Factors.svg`,

TF post-translational modifications

- Phosphorylation, glycosylation, methylation, acetylation, sumoylation, ubiquitination
- Can affect TF activity, stability, localisation, ...

More information: doi:10.1016/j.tips.2013.11.005

Steps in gene expression

1. Recruitment and formation of transcription initiation complex
2. Transcription initiation
3. Elongation
4. Transcription termination
5. Splicing and other RNA processing
6. mRNA transport
7. Translation
8. mRNA degradation

Elongation

- Pol-II moving along DNA, synthesising new pre-mRNA
- Average velocity 2–4 kbp/min
- Pausing
- Falling off

Elongation

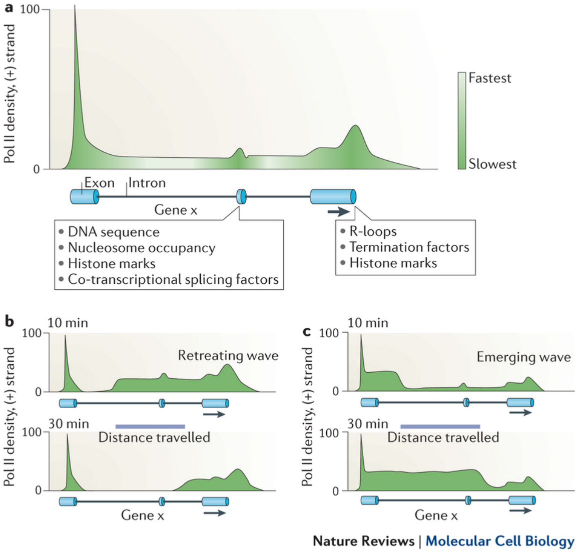


Figure from Jonkers & Lis (2015), doi:10.1038/nrm3953

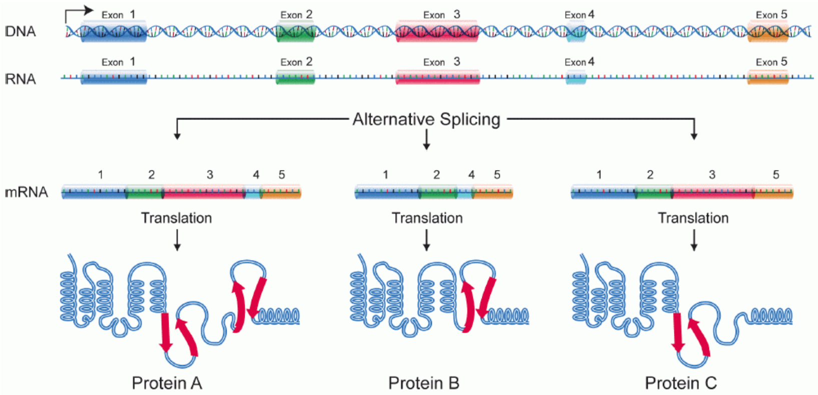
Steps in gene expression

1. Recruitment and formation of transcription initiation complex
2. Transcription initiation
3. Elongation
4. Transcription termination
5. Splicing and other RNA processing
6. mRNA transport
7. Translation
8. mRNA degradation

Splicing

- Often co-transcriptional
- Alternative splicing
- Different intron types (e.g. U12) with different rates
- Coupled with other steps

Splicing

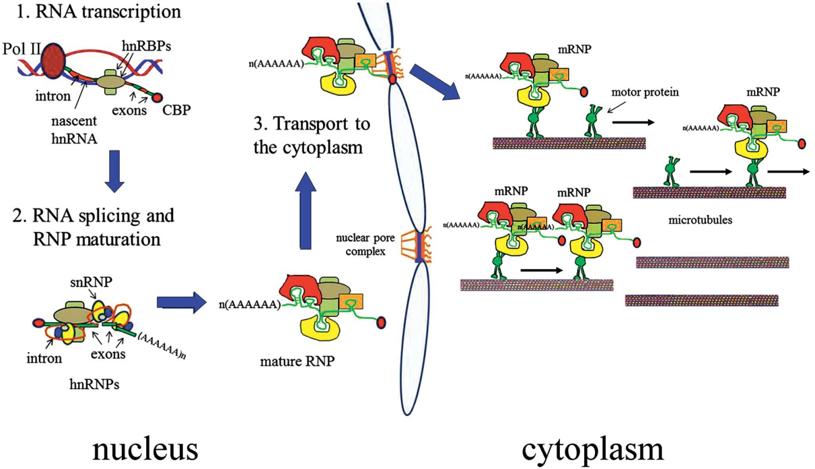


DNA
Exon 1
Exon 2
Exon 3
Exon 4
Exon 5
RNA
Exon 1
Exon 2
Exon 3
Exon 4
Exon 5
Alternative Splicing
mRNA
1 2 3 4 5
1 2 4 5
1 2 3 5
Translation
Translation
Translation
Protein A
Protein B
Protein C

Steps in gene expression

1. Recruitment and formation of transcription initiation complex
2. Transcription initiation
3. Elongation
4. Transcription termination
5. Splicing and other RNA processing
6. mRNA transport
7. Translation
8. mRNA degradation

mRNA transport



Steps in gene expression

1. Recruitment and formation of transcription initiation complex
2. Transcription initiation
3. Elongation
4. Transcription termination
5. Splicing and other RNA processing
6. mRNA transport
7. Translation
8. mRNA degradation

mRNA degradation

- Typical mRNA half lifes:
  - Bacterial: 2-5 min
  - Mammalian: 10 min to more than 10 hours
- No reason for these to be constant, but mechanisms of regulation difficult to study and therefore poorly understood

Regulation by alternative splicing and mRNA degradation

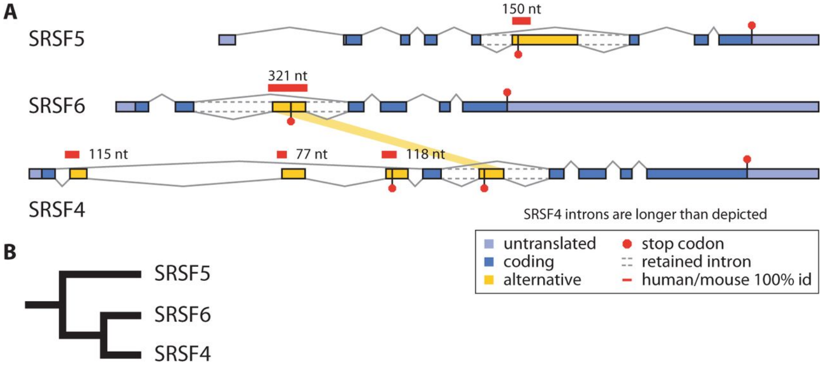


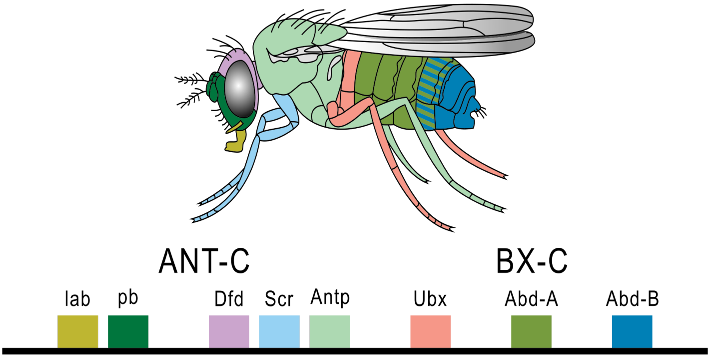
Figure from Lareau & Brenner (2015), doi:10.1093/molbev/msv002, CC-BY

Steps in gene expression

1. Recruitment and formation of transcription initiation complex
2. Transcription initiation
3. Elongation
4. Transcription termination
5. Splicing and other RNA processing
6. mRNA transport
7. Translation
8. mRNA degradation

Regulation of development

Regulation of development



Combinatorial gene regulation by modulation of relative pulse timing

http://www.nature.com/nature/journal/vaop/ncurrent/full/nature15710.html

Outline

Overview of experimental functional genomics

Measurement objectives:

Overview of experimental functional genomics

Measurement objectives:

- Gene expression
- DNA binding locations (TFs, epigenetic marks, RNA polymerase)
- Run-on assays (RNA polymerase, ribosome)
- DNA 3D structure
- Specialised assays (e.g. methylation)

Measurement technologies:

Overview of experimental functional genomics

Measurement objectives:

- Gene expression
- DNA binding locations (TFs, epigenetic marks, RNA polymerase)
- Run-on assays (RNA polymerase, ribosome)
- DNA 3D structure
- Specialised assays (e.g. methylation)

Measurement technologies:

- Microarrays
- Sequencing
- qPCR
- Imaging

Measurement objects:

Overview of experimental functional genomics

Measurement objectives:

- Gene expression
- DNA binding locations (TFs, epigenetic marks, RNA polymerase)
- Run-on assays (RNA polymerase, ribosome)
- DNA 3D structure
- Specialised assays (e.g. methylation)

Measurement technologies:

- Microarrays
- Sequencing
- qPCR
- Imaging

Measurement objects:

- Cell populations
- Single cells

Overview of experimental functional genomics

Measurement objectives:

- Gene expression
- DNA binding locations (TFs, epigenetic marks, RNA polymerase)
- Run-on assays (RNA polymerase, ribosome)
- DNA 3D structure
- Specialised assays (e.g. methylation)

Measurement technologies:

- Microarrays
- Sequencing
- qPCR
- Imaging

Measurement objects:

- Cell populations
- Single cells

Overview of experimental functional genomics

Measurement objectives:

- Gene expression
- DNA binding locations (TFs, epigenetic marks, RNA polymerase)
- Run-on assays (RNA polymerase, ribosome)
- DNA 3D structure
- Specialised assays (e.g. methylation)

Measurement technologies:

- Microarrays
- Sequencing
- qPCR
- Imaging

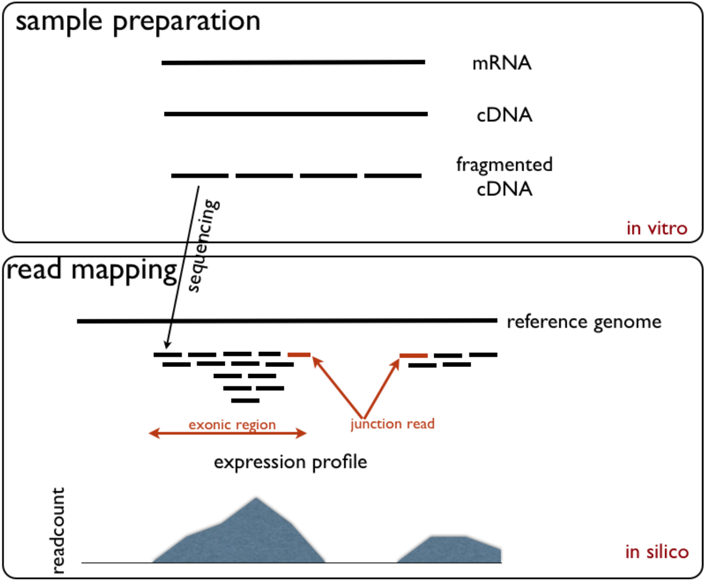
Measurement objects:

- Cell populations
- Single cells

Common data analysis steps

1. Alignment to a reference genome
2. Peak finding (binding measurements)
3. Quantification of features (genes, transcripts, peaks)

Example: RNA-sequencing



Outline

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

Markov jump process:

$$p(X(t + dt)) = f(X(t))$$

Different approaches for simulating a cell

Discrete state – Continuous state

Discrete time – Continuous time

Deterministic – Stochastic

Differential equations:

$$\frac{dX(t)}{dt} = f(X(t))$$

Different approaches for simulating a cell

Discrete state – Continuous state

Discrete time – Continuous time

Deterministic – Stochastic

Stochastic differential equations:

$$dX(t) = f(X(t))dt + g(X(t))dW(t)$$

Outline

Learning goals for this week

- To understand the *objectives* of gene regulation.
- To recognise *basic mechanisms* of 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:

- Inferring regulatory networks

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:

1. Read Gillespie (1977), especially Secs. I, IIIB, IIIC
2. Read Gillespie (2000), especially Secs. I, II, III
3. Read Gillespie (2000), especially Secs. I, II, IV