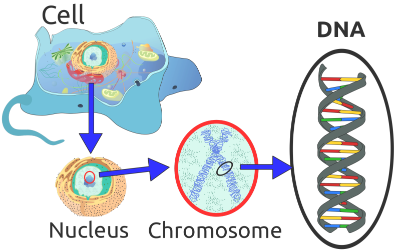
RNA Velocity: a mathematical model to predict cellular differentiations

Loïc Demeulenaere

Comprehensible Seminars – Institute of Mathematics, ULiège

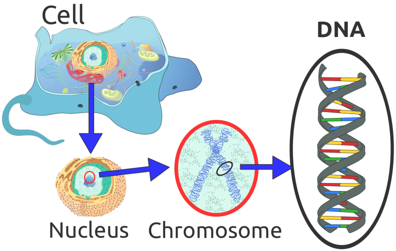
June 26, 2019

Deoxyribonucleic acid (DNA)



([5])

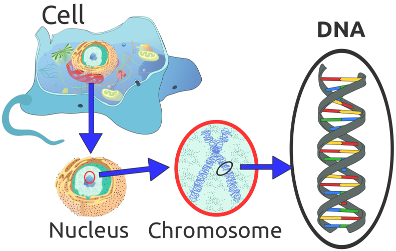
Deoxyribonucleic acid (DNA)



([5])

- Basic units: *nucleotides* (sugar (=deoxyribose) + phosphate + nucleic base)
  4 types: Adenine (A), Cytosine (C), Guanine (G), Thymine (T)

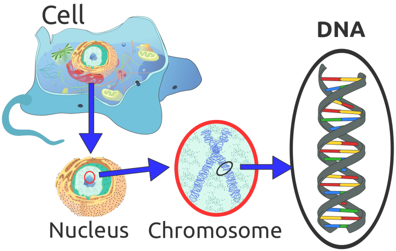
Deoxyribonucleic acid (DNA)



([5])

- Basic units: *nucleotides* (sugar (=deoxyribose) + phosphate + nucleic base)
  4 types: Adenine (A), Cytosine (C), Guanine (G), Thymine (T)
- One DNA strand: concatenation of millions of nucleotides (on average 140×10^6 nucleotides, 4.8 cm) ([3])

Deoxyribonucleic acid (DNA)



([5])

- Basic units: *nucleotides* (sugar (=deoxyribose) + phosphate + nucleic base)
  4 types: Adenine (A), Cytosine (C), Guanine (G), Thymine (T)
- One DNA strand: concatenation of millions of nucleotides (on average 140×10^6 nucleotides, 4.8 cm) ([3])
- Structure:
  - Double-helix (2 strands, with links A-T or C-G)
  - Folded to form *Chromosomes* (23 pairs for human beings)

Genes and ribonucleic acid (RNA)

- *Genes*: fraction of DNA molecules leading to the synthesis of *RNA* or *proteins*

Genes and ribonucleic acid (RNA)

- *Genes*: fraction of DNA molecules leading to the synthesis of *RNA* or *proteins*
- Genes determine traits of an organism (colour of hair or eyes, height, etc.)

Genes and ribonucleic acid (RNA)

- *Genes*: fraction of DNA molecules leading to the synthesis of *RNA* or *proteins*
- Genes determine traits of an organism (colour of hair or eyes, height, etc.)

Caution!

- All the genes are present in every cell, but are not expressed in every cell!

Genes and ribonucleic acid (RNA)

- *Genes*: fraction of DNA molecules leading to the synthesis of *RNA* or *proteins*
- Genes determine traits of an organism (colour of hair or eyes, height, etc.)

Caution!

- All the genes are present in every cell, but are not expressed in every cell!
- Genetic activities → synthesis of (messenger) RNA

Genes and ribonucleic acid (RNA)

- *Genes*: fraction of DNA molecules leading to the synthesis of *RNA* or *proteins*

- Genes determine traits of an organism (colour of hair or eyes, height, etc.)

Caution!

- All the genes are present in every cell, but are not expressed in every cell!
- Genetic activities $\rightarrow$ synthesis of (messenger) RNA

- Ribonucleic Acid (RNA): 4 nucleotides (A,C,G, and *Uracil* (U)), with ribose as sugar (usually, one simple strand)

Synthesis of RNA

1. *Transcription*: copy of a gene → production of *precursor* RNA

Synthesis of RNA

1. *Transcription*: copy of a gene → production of *precursor* RNA
2. *Splicing*: maturation, excision of certain fragments of the precursor RNA → production of mature (i.e. functional) RNA

Synthesis of RNA

1. *Transcription*: copy of a gene → production of *precursor* RNA
2. *Splicing*: maturation, excision of certain fragments of the precursor RNA → production of mature (i.e. functional) RNA

Coding genes and mRNA

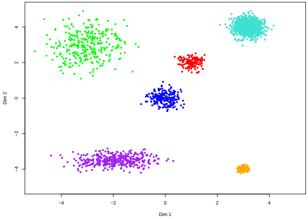
- Among genes, there exist *coding genes*: they produce *messenger RNA* (mRNA), which is used to build some proteins needed by the organism
- Other genes are called *RNA genes*

Purpose

- Usually, in genomics: clustering of cells
  $\rightsquigarrow$ study of *differentiated* cells...

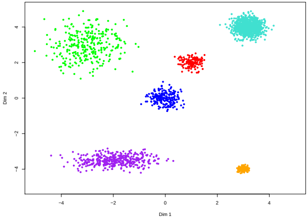
Purpose

- Usually, in genomics: clustering of cells
  $\rightsquigarrow$ study of *differentiated* cells...



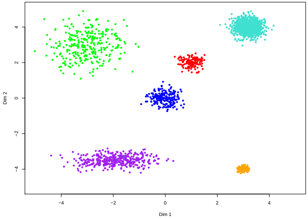
Purpose

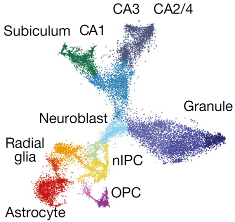
- Usually, in genomics: clustering of cells
  $\rightsquigarrow$ study of *differentiated* cells...
- Other point of view: (continuous) differentiation of cells (from stem/precursor cells)



Purpose

- Usually, in genomics: clustering of cells
  $\rightsquigarrow$ study of *differentiated* cells...
- Other point of view: (continuous) differentiation of cells (from stem/precursor cells)



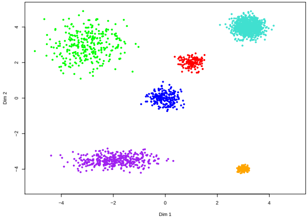


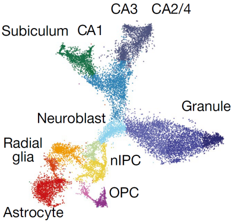
([1])

Purpose

- Usually, in genomics: clustering of cells
  $\rightsquigarrow$ study of *differentiated* cells...

- Other point of view: (continuous) differentiation of cells (from stem/precursor cells)



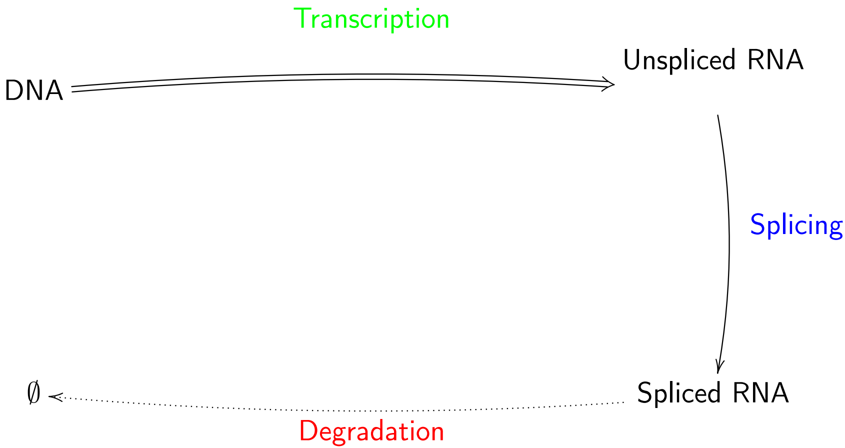


([1])

- Purpose: **modelling this differentiation by study "RNA dynamics"** (transcription + splicing)

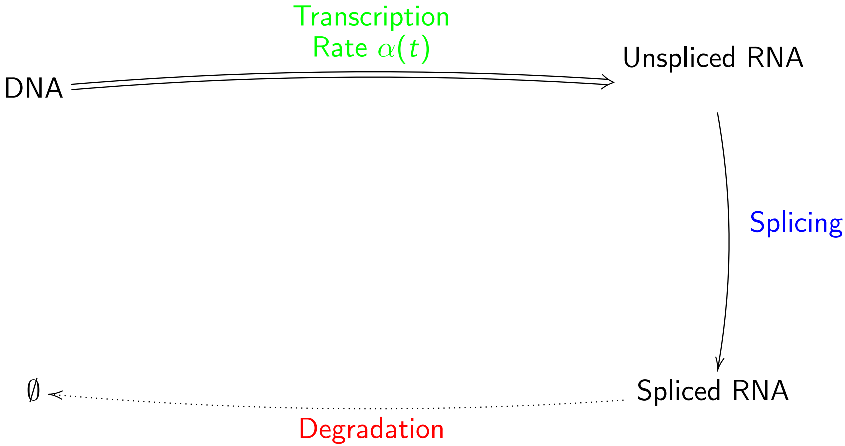
RNA dynamics

In **one cell**, for **one gene** ([1])...



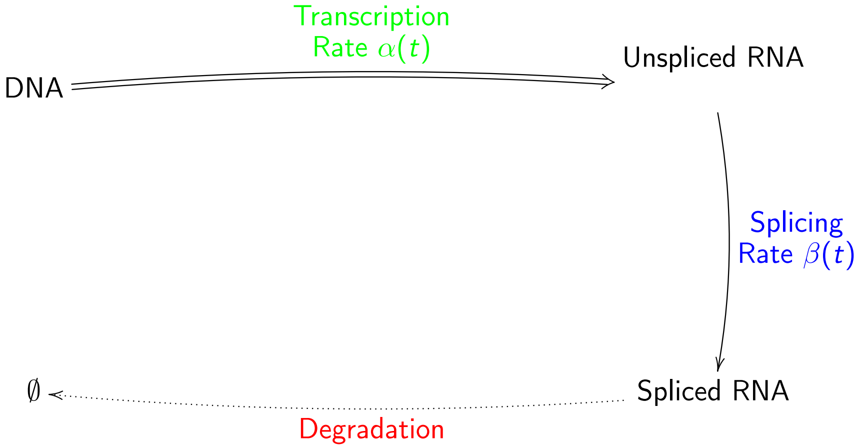
RNA dynamics

In **one cell**, for **one gene** ([1])...



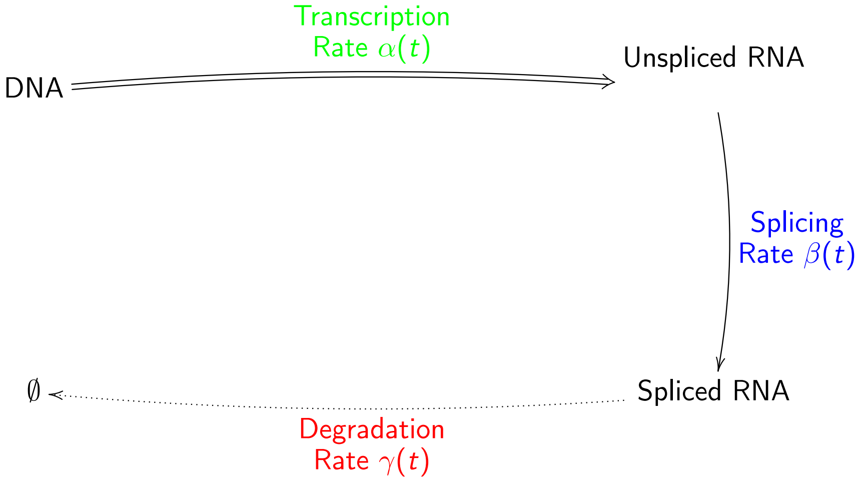
RNA dynamics

In **one cell**, for **one gene** ([1])...



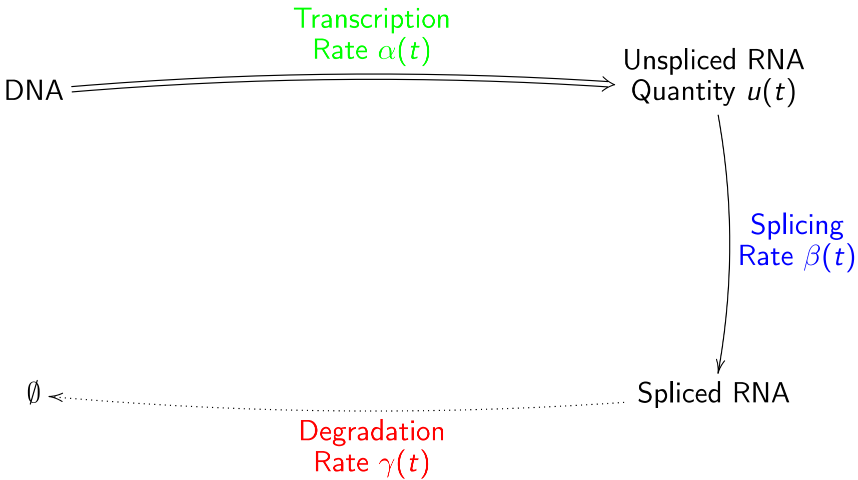
RNA dynamics

In **one** **cell**, for **one** **gene** ([1])...



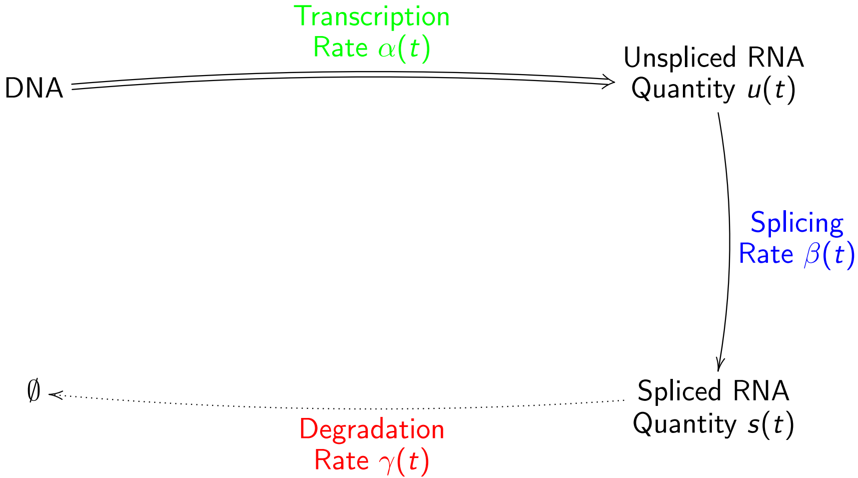
RNA dynamics

In **one cell**, for **one gene** ([1])...



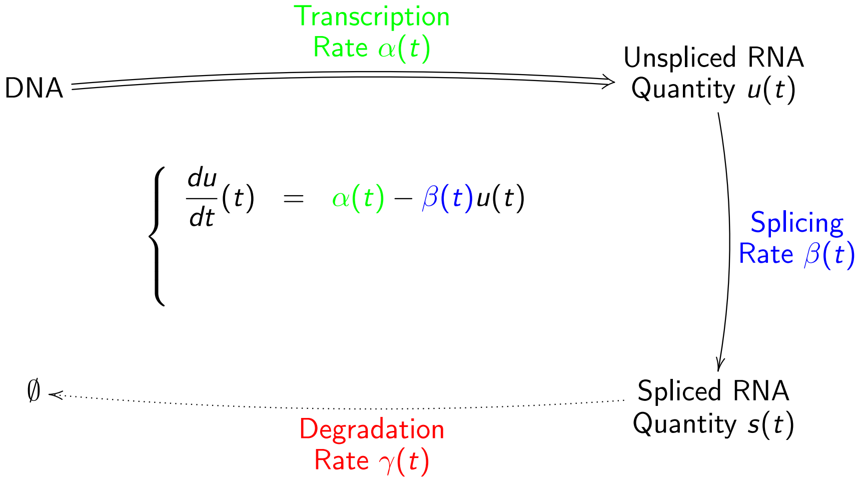
RNA dynamics

In __one__ **cell**, for __one__ **gene** ([1])...



RNA dynamics

In **one cell**, for **one gene** ([1])...



$$\left\{ \frac{du}{dt}(t) = \alpha(t) - \beta(t)u(t) \right.$$

RNA dynamics

In **<u>one</u> cell**, for **<u>one</u> gene** ([1])...

Transcription
Rate $\alpha(t)$

DNA ⟹ Unspliced RNA
Quantity $u(t)$

$$\begin{cases} \frac{du}{dt}(t) & = & \alpha(t) - \beta(t)u(t) \\ \frac{ds}{dt}(t) & = & \beta(t)u(t) - \gamma(t)s(t) \end{cases}$$

Splicing
Rate $\beta(t)$

Spliced RNA
Quantity $s(t)$

$\emptyset$

Degradation
Rate $\gamma(t)$

Remarks

1. Parameters
   - Assumption/simplification: all the parameters are constant and

$$\alpha \geq 0, \qquad \beta > 0, \qquad \gamma > 0$$

Remarks

1. Parameters
   - Assumption/simplification: all the parameters are constant and

$$\alpha \geq 0, \qquad \beta > 0, \qquad \gamma > 0$$

   - Change of unit of time s.t. $\beta = 1$

Remarks

1. Parameters
   - Assumption/simplification: all the parameters are constant and

   $$\alpha \geq 0, \qquad \beta > 0, \qquad \gamma > 0$$

   - Change of unit of time s.t. $\beta = 1$

2. Distribution
   - $u(t)$ and $s(t)$ are continuous functions describing integer numbers: **expected values**!

Remarks

1. Parameters
    - Assumption/simplification: all the parameters are constant and

$$\alpha \geq 0, \qquad \beta > 0, \qquad \gamma > 0$$

    - Change of unit of time s.t. $\beta = 1$

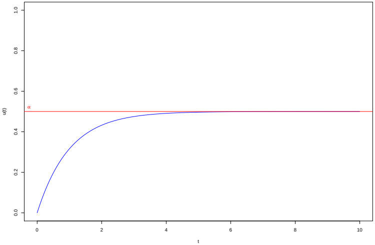
2. Distribution
    - $u(t)$ and $s(t)$ are continuous functions describing integer numbers: **expected values**!
    - **Real** integer values: (asymptotic) 2-dimensional Poisson distribution

Solution of the first equation

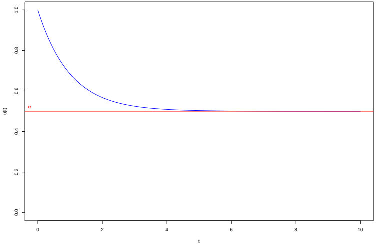
If $u_0 := u(0)$,

$$u(t) = \alpha + (u_0 - \alpha)e^{-t}.$$

$u_0 < \alpha$



$u_0 > \alpha$



In all cases,

$$\lim_{t \to \infty} u(t) = \alpha.$$

Solution of the second equation

If $u_0 := u(0)$ and $s_0 := s(0)$,

$$s(t) = \begin{cases} \dfrac{\alpha}{\gamma} + \dfrac{u_0 - \alpha}{\gamma - 1} e^{-t} + \left(s_0 + \dfrac{\alpha - u_0}{\gamma - 1} - \dfrac{\alpha}{\gamma} \right) e^{-\gamma t} & \text{if} \quad \gamma \neq 1 \\ \alpha + \left[(u_0 - \alpha) t + s_0 - \alpha \right] e^{-t} & \text{if} \quad \gamma = 1 (= \beta). \end{cases}$$

Solution of the second equation

If $u_0 := u(0)$ and $s_0 := s(0)$,

$$s(t) = \begin{cases} \dfrac{\alpha}{\gamma} + \dfrac{u_0 - \alpha}{\gamma - 1} e^{-t} + \left(s_0 + \dfrac{\alpha - u_0}{\gamma - 1} - \dfrac{\alpha}{\gamma} \right) e^{-\gamma t} & \text{if} \quad \gamma \neq 1 \\ \alpha + \left[(u_0 - \alpha) t + s_0 - \alpha \right] e^{-t} & \text{if} \quad \gamma = 1 (= \beta). \end{cases}$$

We always have

$$\lim_{t \to \infty} s(t) = \frac{\alpha}{\gamma}$$

Solution of the second equation

If $u_0 := u(0)$ and $s_0 := s(0)$,

$$s(t) = \begin{cases} \frac{\alpha}{\gamma} + \frac{u_0 - \alpha}{\gamma - 1} e^{-t} + \left(s_0 + \frac{\alpha - u_0}{\gamma - 1} - \frac{\alpha}{\gamma} \right) e^{-\gamma t} & \text{if} \quad \gamma \neq 1 \\ \alpha + \left[(u_0 - \alpha) t + s_0 - \alpha \right] e^{-t} & \text{if} \quad \gamma = 1 (= \beta). \end{cases}$$

We always have

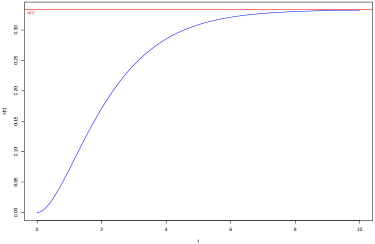
$$\lim_{t \to \infty} s(t) = \frac{\alpha}{\gamma}$$

and so

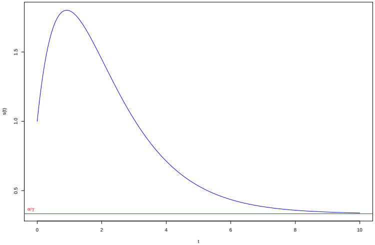
$$\boxed{\lim_{t \to \infty} \frac{u(t)}{s(t)} = \gamma.}$$

Solution of the second equation

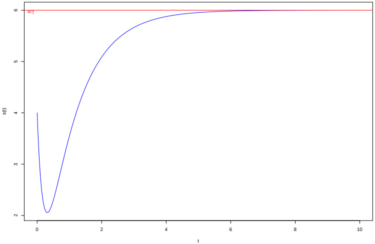
$u_0 = s_0 = 0, \ \alpha = 0.25, \ \gamma = 0.75$



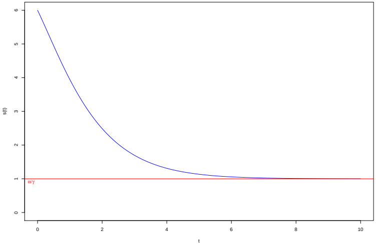
$u_0 = 3, \ s_0 = 1, \ \alpha = 0.25, \ \gamma = 0.75$



$u_0 = 3, \ s_0 = 4, \ \alpha = 30, \ \gamma = 5$



$u_0 = 4, \ s_0 = 6, \ \alpha = 1, \ \gamma = 1$



RNA Velocity

In **one** **cell**, for p **genes**... ([1])

RNA Velocity

In **one cell**, for p **genes**... ([1])

- $s_j(t)$ quantity of spliced RNA associated to the j^{th} gene (at time t)

RNA Velocity

In **one** **cell**, for p **genes**... ([1])

- $s_j(t)$ quantity of spliced RNA associated to the j^{th} gene (at time t)
- s_j has its **own parameters** $\alpha_j \geq 0$, $\beta_j = 1$, and $\gamma_j > 0$

RNA Velocity

In <u>**one**</u> **cell**, for p **genes**... ([1])

- $s_j(t)$ quantity of spliced RNA associated to the j^{th} gene (at time t)
- s_j has its **own parameters** $\alpha_j \geq 0$, $\beta_j = 1$, and $\gamma_j > 0$

Caution!

$\beta_j = 1$ for all j: the rates of splicing are equal for all genes!

RNA Velocity

In **one** **cell**, for p **genes**... ([1])

- $s_j(t)$ quantity of spliced RNA associated to the j^{th} gene (at time t)
- s_j has its **own parameters** $\alpha_j \geq 0$, $\beta_j = 1$, and $\gamma_j > 0$

  Caution!

  $\beta_j = 1$ for all j: the rates of splicing are equal for all genes!

- The cell is "characterized" (at time t) by

$$\vec{s}(t) = (s_1(t), \cdots, s_p(t))$$

  in a p-dimensional space (space of spliced quantities)

RNA Velocity

In **one** **cell**, for p **genes**... ([1])

- $s_j(t)$ quantity of spliced RNA associated to the j^{th} gene (at time t)
- s_j has its **own parameters** $\alpha_j \geq 0$, $\beta_j = 1$, and $\gamma_j > 0$

  Caution!

  $\beta_j = 1$ for all j: the rates of splicing are equal for all genes!
- The cell is "characterized" (at time t) by

$$\vec{s}(t) = (s_1(t), \cdots, s_p(t))$$

  in a p-dimensional space (space of spliced quantities)

Definition

The *RNA velocity* of the cell is

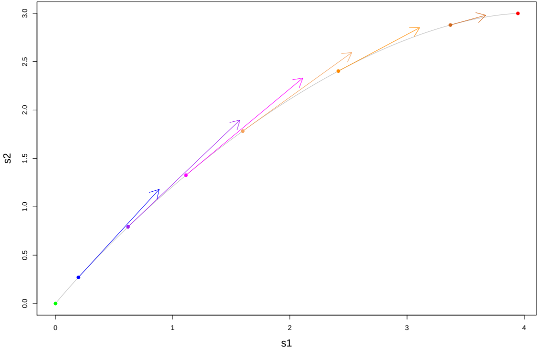
$$\vec{v} := \frac{d\vec{s}}{dt} = \left(\frac{ds_1}{dt}, \cdots, \frac{ds_p}{dt} \right).$$

Representations

Help for interpretation...

Example ($p = 2$)

$\alpha_1 = 2$, $\gamma_1 = 0.5$; $\alpha_2 = 3$, $\gamma_2 = 1$



- Grey curve: trajectory of the cell in the space of spliced quantities $(t \mapsto (s_1(t), s_2(t)))$
- Vectors: RNA velocities
- Red point: "steady" state

Representations

For $p > 3$?

We need some *dimensional reductions*!

Representations

For $p > 3$?

We need some *dimensional reductions*!

- Principle component analysis

Representations

For $p > 3$?

We need some *dimensional reductions*!

- Principle component analysis: quite natural, projection on P.C.;

Representations

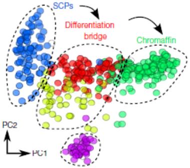
For $p > 3$?

We need some *dimensional reductions*!

- Principle component analysis: quite natural, projection on P.C.;
- There exist some non-linear techniques (t-SNE ([4]), UMAP ([2])). Idea: arrows joining cells aligned along velocities...

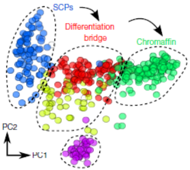
Representations

Example: Schwann cell precursors (coming from [1])

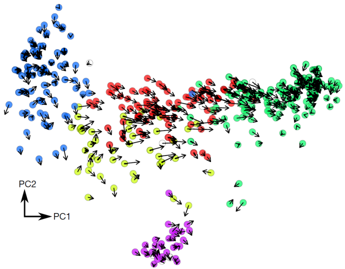


Representations

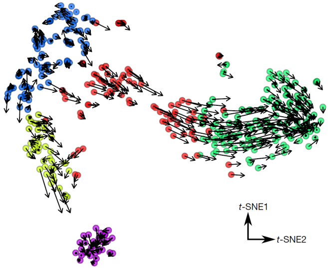
Example: Schwann cell precursors (coming from [1])



PCA



t-SNE



Estimation of γ

For <u>**one**</u> **fixed gene** and a given **sample of cells** (of size n)...

Assumption

Degradation coefficient γ depends **on the gene**, but <u>**not**</u> **on the cell**!

Estimation of γ

For <u>**one**</u> **fixed gene** and a given **sample of cells** (of size n)...

Assumption

Degradation coefficient γ depends **on the gene**, but <u>**not**</u> **on the cell**!
According to [1], 89 % of genes respect this property (filters for others).

Estimation of γ

For **one fixed gene** and a given **sample of cells** (of size n)...

Assumption

Degradation coefficient γ depends **on the gene**, but **not on the cell**!
According to [1], 89 % of genes respect this property (filters for others).

How can we estimate the parameter γ associated to the fixed gene?

Phase portrait

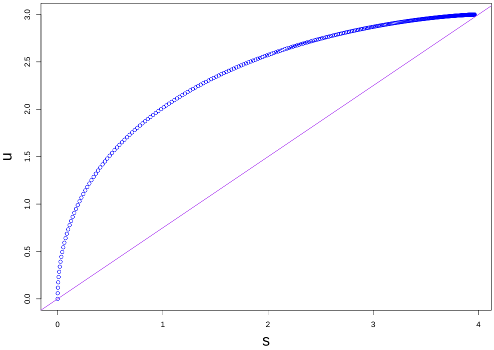
- Reminder: in every cell, $\lim_{t\to\infty}(u(t)/s(t)) = \gamma$ so $u(t) \approx \gamma s(t)$ if $t \gg 0$.

Phase portrait

- Reminder: in every cell, $\lim_{t\to\infty}(u(t)/s(t)) = \gamma$ so $u(t) \approx \gamma s(t)$ if $t \gg 0$.
- Phase portrait: graphic s *vs* u

Theoritical example ($\alpha = 3,\ \gamma = 0.75$)

400 cells, uniformly generated in time



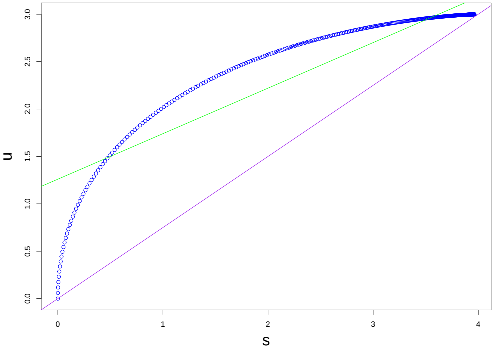
- Straight line $u = \gamma s$

Phase portrait

- Reminder: in every cell, $\lim_{t\to\infty}(u(t)/s(t)) = \gamma$ so $u(t) \approx \gamma s(t)$ if $t \gg 0$.
- Phase portrait: graphic s *vs* u

Theoritical example ($\alpha = 3,\ \gamma = 0.75$)

400 cells, uniformly generated in time



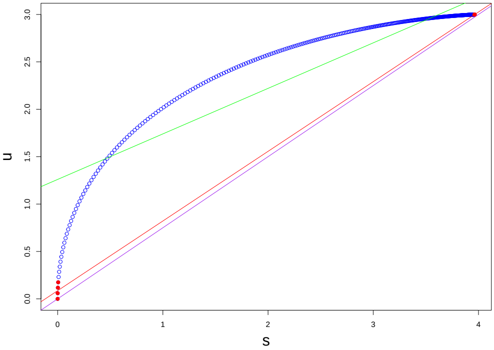
- Straight line $u = \gamma s$
- Linear regression: $\gamma \approx 0.4803$

Phase portrait

- Reminder: in every cell, $\lim_{t\to\infty}(u(t)/s(t)) = \gamma$ so $u(t) \approx \gamma s(t)$ if $t \gg 0$.
- Phase portrait: graphic s *vs* u

Theoritical example ($\alpha = 3,\ \gamma = 0.75$)

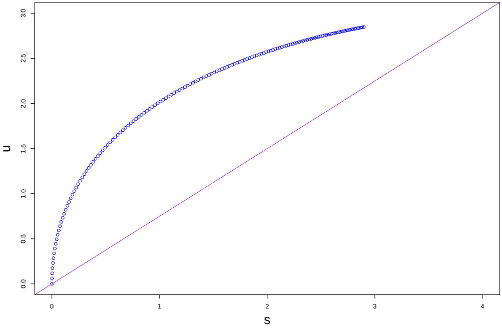
400 cells, uniformly generated in time



- Straight line $u = \gamma s$
- Linear regression: $\gamma \approx 0.4803$
- Extreme-quantile linear regression (1%): $\gamma \approx 0.73493$

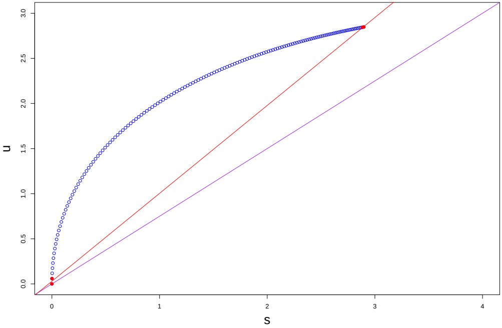
Estimation of γ: caution!

300 from the 400 previous cells...



Estimation of γ: caution!

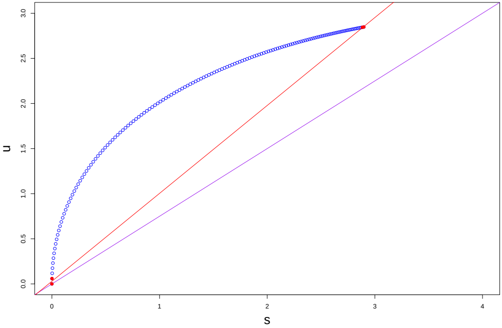
300 from the 400 previous cells...



$\gamma \approx 0.97433$

Estimation of γ: caution!

300 from the 400 previous cells...



$\gamma \approx 0.97433$

Necessary condition: as many differentiation steps as possible...

Estimation of α and short-term approximation

- Estimation of α: Difficult, α depends on **genes AND cells**...

Estimation of α and short-term approximation

- Estimation of α: Difficult, α depends on **genes AND cells**...
- Idea: short-term approximations (one cell, one gene)

Estimation of α and short-term approximation

- Estimation of α: Difficult, α depends on **genes AND cells**...

- Idea: short-term approximations (one cell, one gene):
  - **Model I**: $v := \frac{ds}{dt}$ is $\pm$ constant:

$$s(t) = vt + s_0,$$

with $v := u_0 - \gamma s_0$.

Estimation of α and short-term approximation

- Estimation of α: Difficult, α depends on **genes AND cells**...
- Idea: short-term approximations (one cell, one gene):
  - **Model I**: $v := \frac{ds}{dt}$ is $\pm$ constant:

$$s(t) = vt + s_0,$$

    with $v := u_0 - \gamma s_0$.

  - **Model II** u is $\pm$ constant:

$$s(t) = \frac{u_0}{\gamma} + \left(s_0 - \frac{u_0}{\gamma}\right) e^{-\gamma t}.$$

Long-term prediction

Long-term prediction

Use of a Markov Model

Idea: flow of cells through manifold, following RNA velocities along close neighbours...

Research project

- Using RNA Velocity to detect genes significantly responsible for cellular differentiations

Research project

- Using RNA Velocity to detect genes significantly responsible for cellular differentiations

- 1^{st} application: study of stem cells in retina cells and their differentiated states

Research project

- Using RNA Velocity to detect genes significantly responsible for cellular differentiations
- 1^{st} application: study of stem cells in retina cells and their differentiated states
- 2^{nd} application: study of a genetic disease ("S-cone enhanced Syndrome") and detection of genetic differences between healthy and ill patients

Thank you for your attention!

References I

G. La Manno et al.
RNA velocity of single cells.
Nature, 560:494–516, 2018.

L. McInnes, J. Healy, and J. Melville.
UMAP: Uniform Manifold Approximation and Projection for Dimension Reduction.
ArXiv e-prints, 2018.

T. Strachan, J. Goodship, and P. Chinnery.
Genetics and genomics in medicine.
Garland Science, New York, 2015.

L. Van der Maaten and G. Hinton.
Visualizing Data using t-SNE.
Journal of Machine Learning Research, 9:2579–2605, 2008.

References II

Wikipedia.
DNA.
Accessed June 12, 2019.