Making a computer out of junk DNA

Josh Deutsch
UCSC

Funded by the Foundational Questions Institute
http://fqxi.org.

Outline

Outline

- Overview of development

Outline

- Overview of development
  - Cell signaling

Outline

- Overview of development
  - Cell signaling
  - Gene regulation

Outline

- Overview of development
  - Cell signaling
  - Gene regulation
- Gene networks

Outline

- Overview of development
  - Cell signaling
  - Gene regulation
- Gene networks
- Junk DNA

Outline

- Overview of development
  - Cell signaling
  - Gene regulation
- Gene networks
- Junk DNA
  - Basic Physics (back of the envelope)

Outline

- Overview of development
  - Cell signaling
  - Gene regulation
- Gene networks
- Junk DNA
  - Basic Physics (back of the envelope)
  - Equilibration, degradation, creation

Outline

- Overview of development
  - Cell signaling
  - Gene regulation
- Gene networks
- Junk DNA
  - Basic Physics (back of the envelope)
  - Equilibration, degradation, creation
- Neural networks

Outline

- Overview of development
  - Cell signaling
  - Gene regulation
- Gene networks
- Junk DNA
  - Basic Physics (back of the envelope)
  - Equilibration, degradation, creation
- Neural networks
  - Boltzmann Machine/Hopfield model

Outline

- Overview of development
  - Cell signaling
  - Gene regulation
- Gene networks
- Junk DNA
  - Basic Physics (back of the envelope)
  - Equilibration, degradation, creation
- Neural networks
  - Boltzmann Machine/Hopfield model
- Junk RNA <-> Neural Network

Outline

- Overview of development
  - Cell signaling
  - Gene regulation
- Gene networks
- Junk DNA
  - Basic Physics (back of the envelope)
  - Equilibration, degradation, creation
- Neural networks
  - Boltzmann Machine/Hopfield model
- Junk RNA <-> Neural Network
- Plausibility

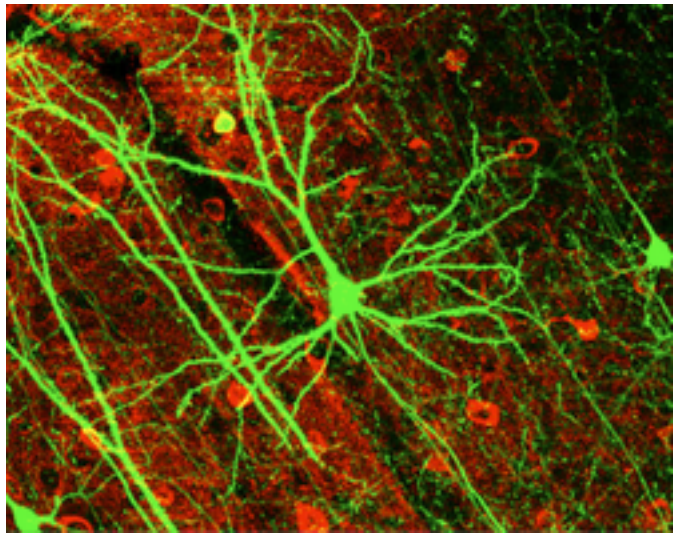
Outline

- Overview of development
  - Cell signaling
  - Gene regulation
- Gene networks
- Junk DNA
  - Basic Physics (back of the envelope)
  - Equilibration, degradation, creation
- Neural networks
  - Boltzmann Machine/Hopfield model
- Junk RNA <-> Neural Network
- Plausibility
- Outlook

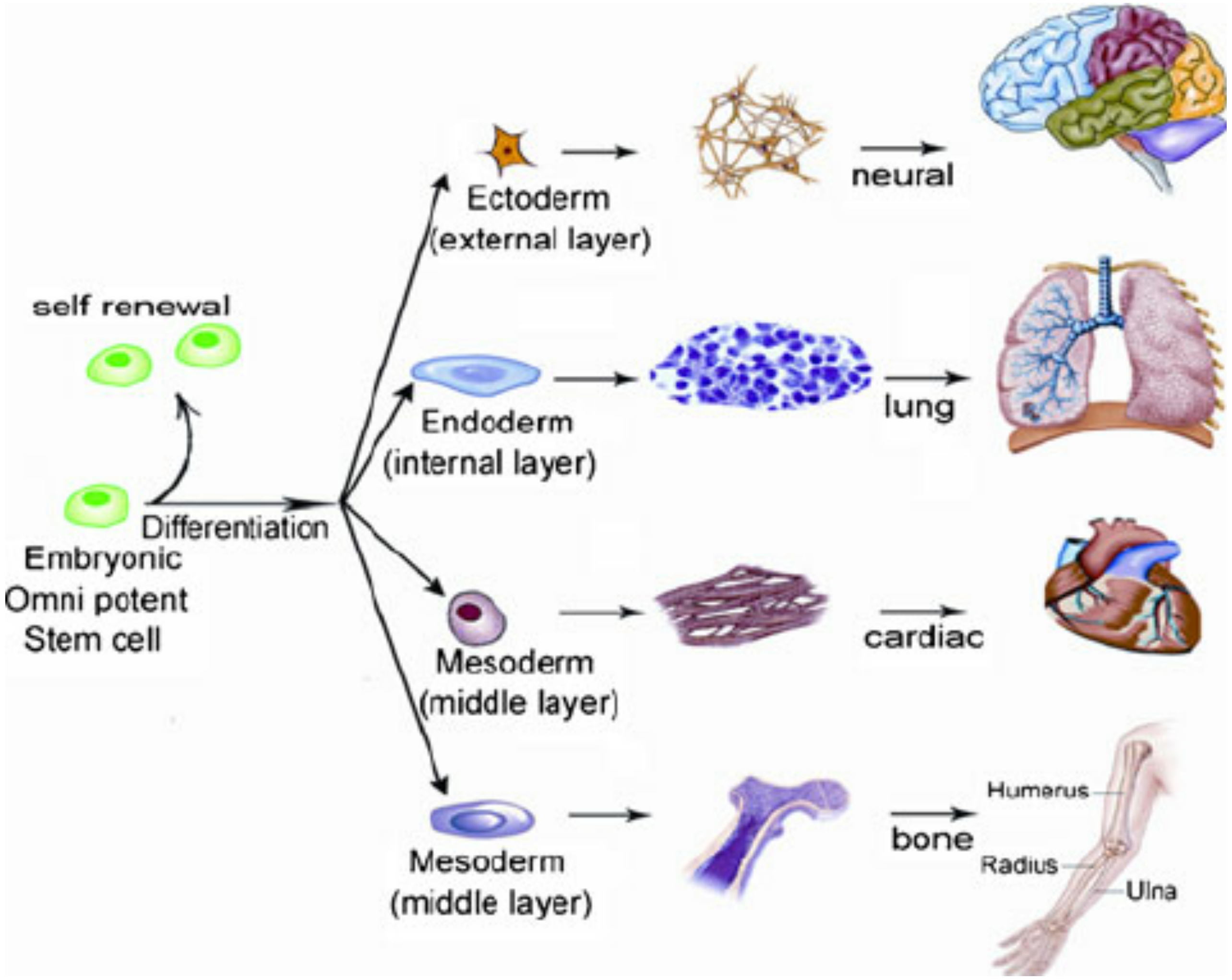
Development

How does an organism develop from a single cell?

How does a brain develop?



Cells receive signals from neighboring cells and their environment, causing them to differentiate.



Cell signaling

Cell signaling

- Many mechanisms exist to transfer information to the cell.

Cell signaling

- Many mechanisms exist to transfer information to the cell.

- This information can get communicated to the cell nucleus causing it to change its state.

Cell signaling

- Many mechanisms exist to transfer information to the cell.

- This information can get communicated to the cell nucleus causing it to change its state.

- This changes what the cell does, like what proteins it will produce.

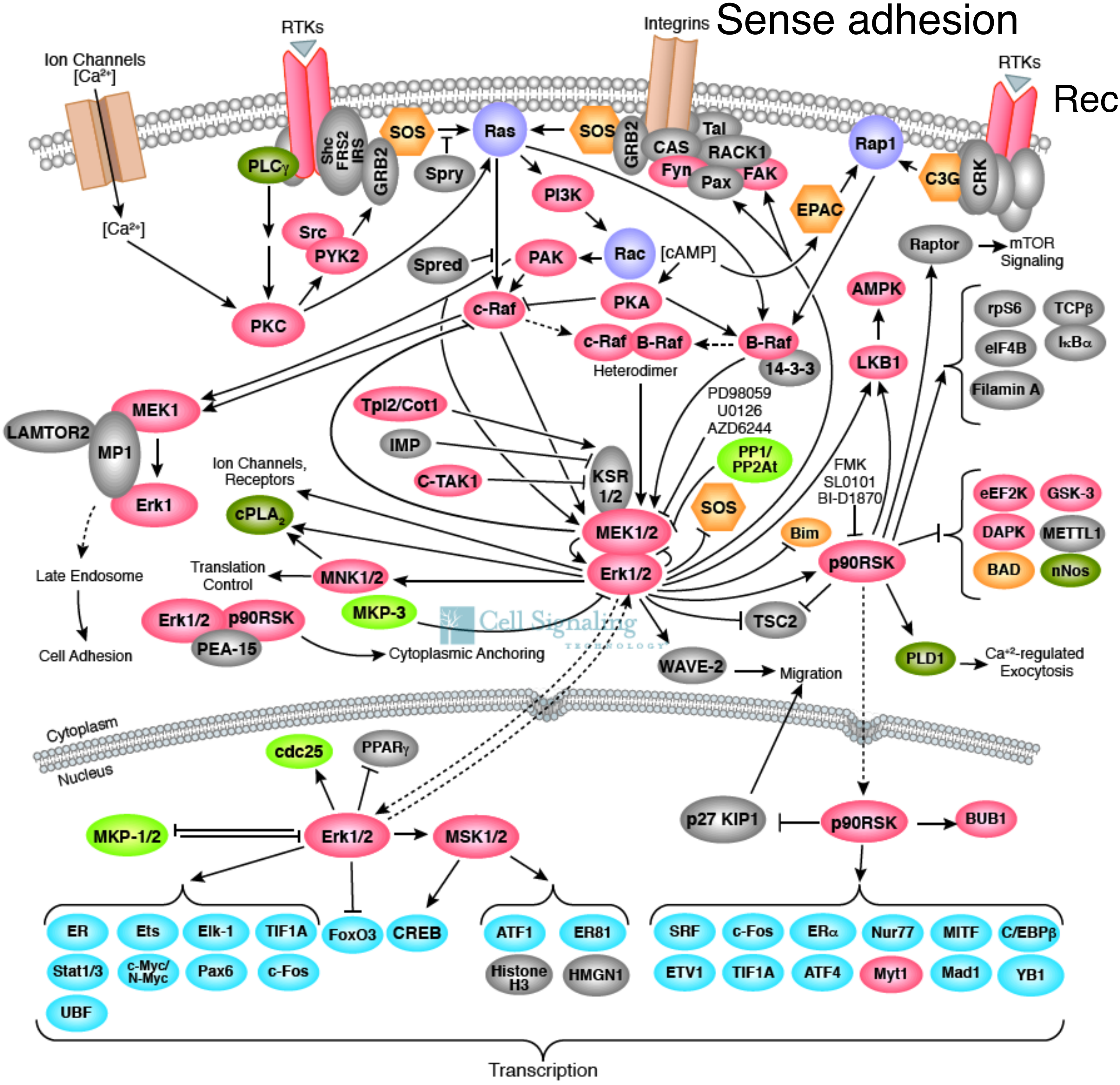
Cell signaling

- Many mechanisms exist to transfer information to the cell.

- This information can get communicated to the cell nucleus causing it to change its state.

- This changes what the cell does, like what proteins it will produce.

- It can tell it for example, to stop growing or form a synapse with another cell.



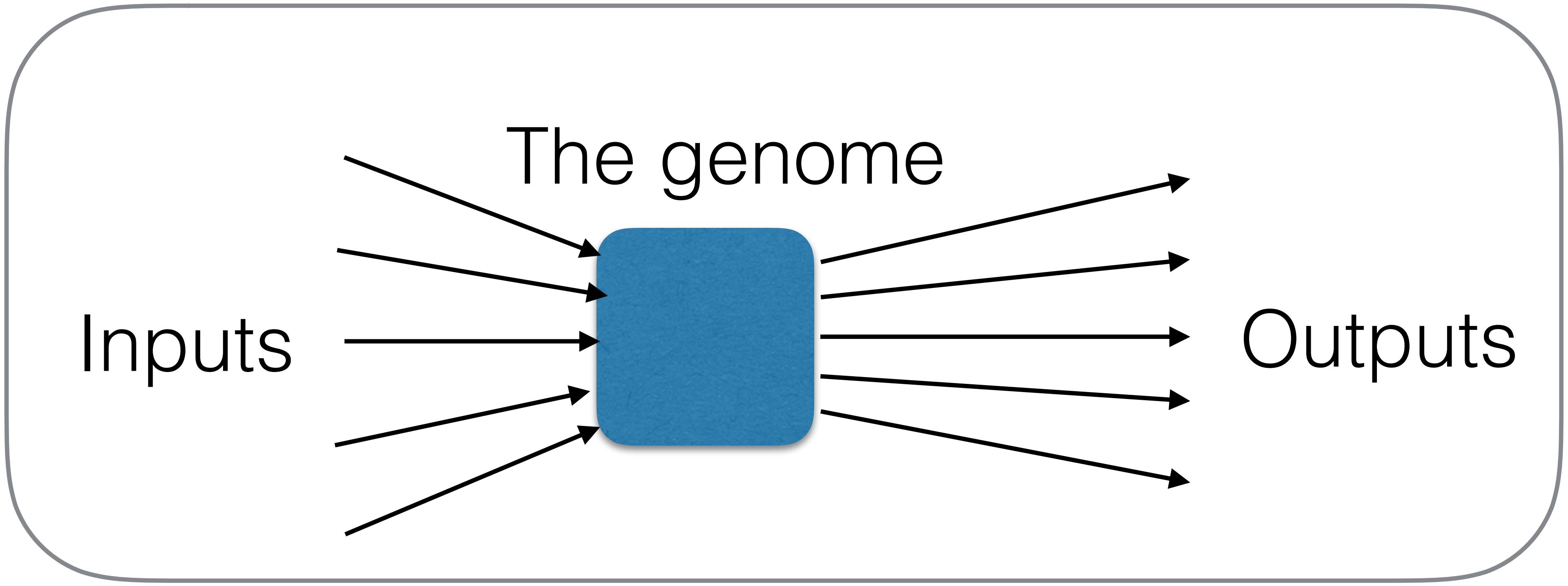
Sense adhesion

Receptor Tyrosine Kinase

Proteins involved in the MAPK/Erk pathway

Communicates surface receptors with DNA

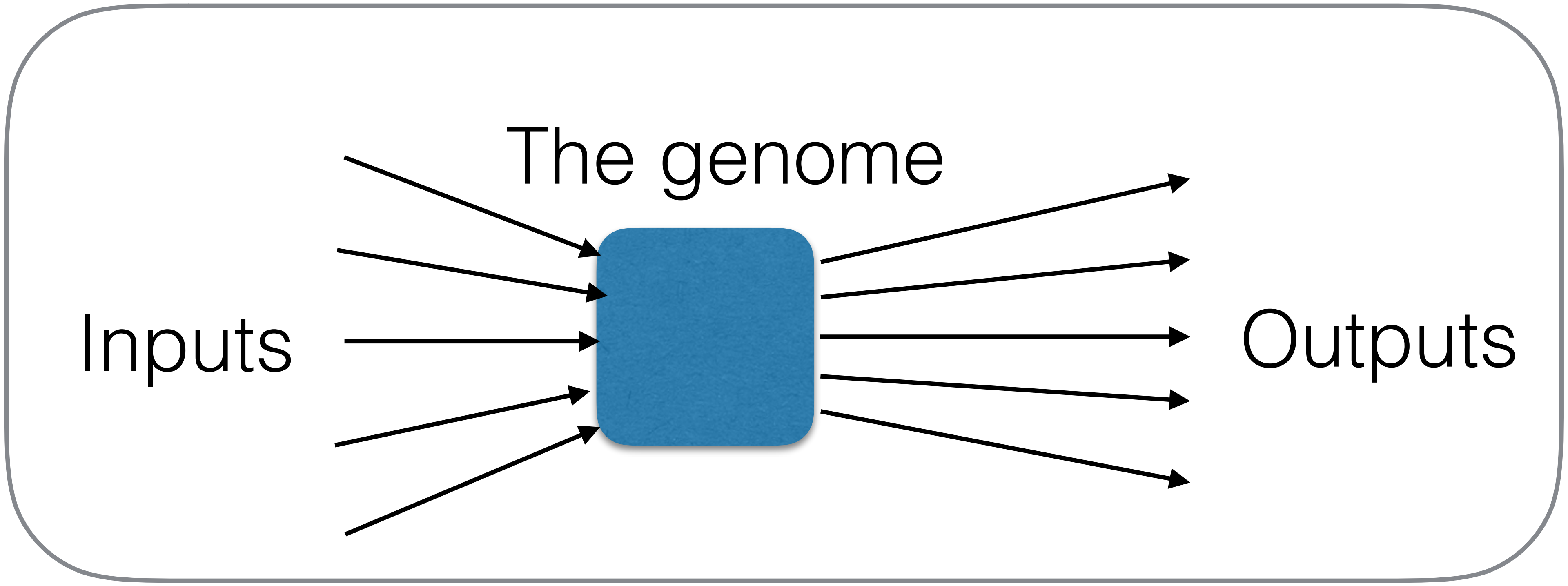
Gene regulation



A cell

Gene regulation

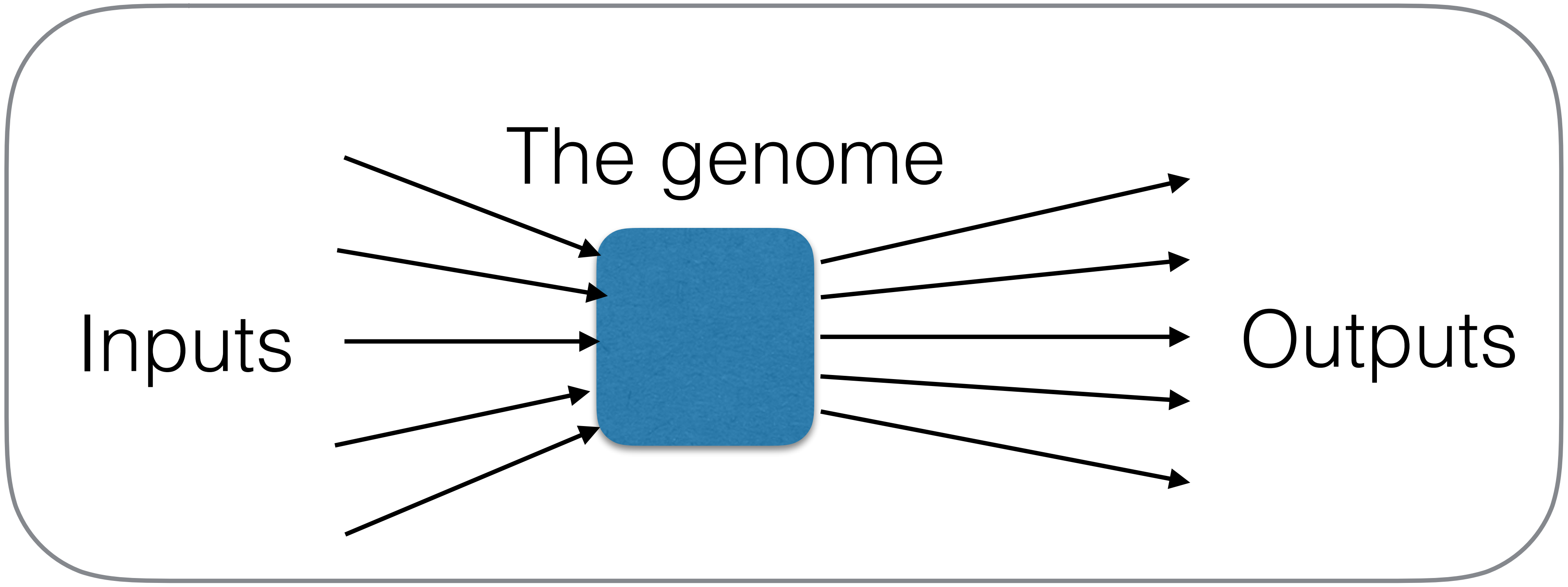
- The input proteins modify the state of the genome to produce output proteins.



A cell

Gene regulation

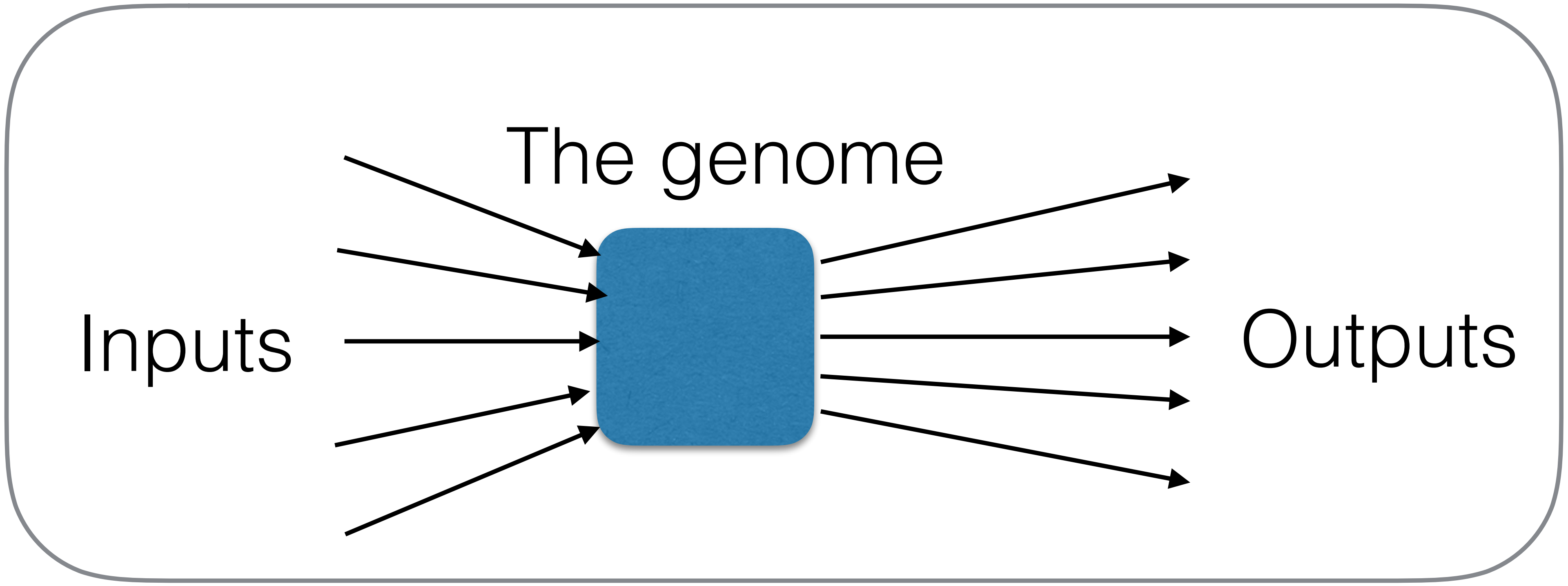
- The input proteins modify the state of the genome to produce output proteins.
- The relationship between these two is extremely complex.



A cell

Gene regulation

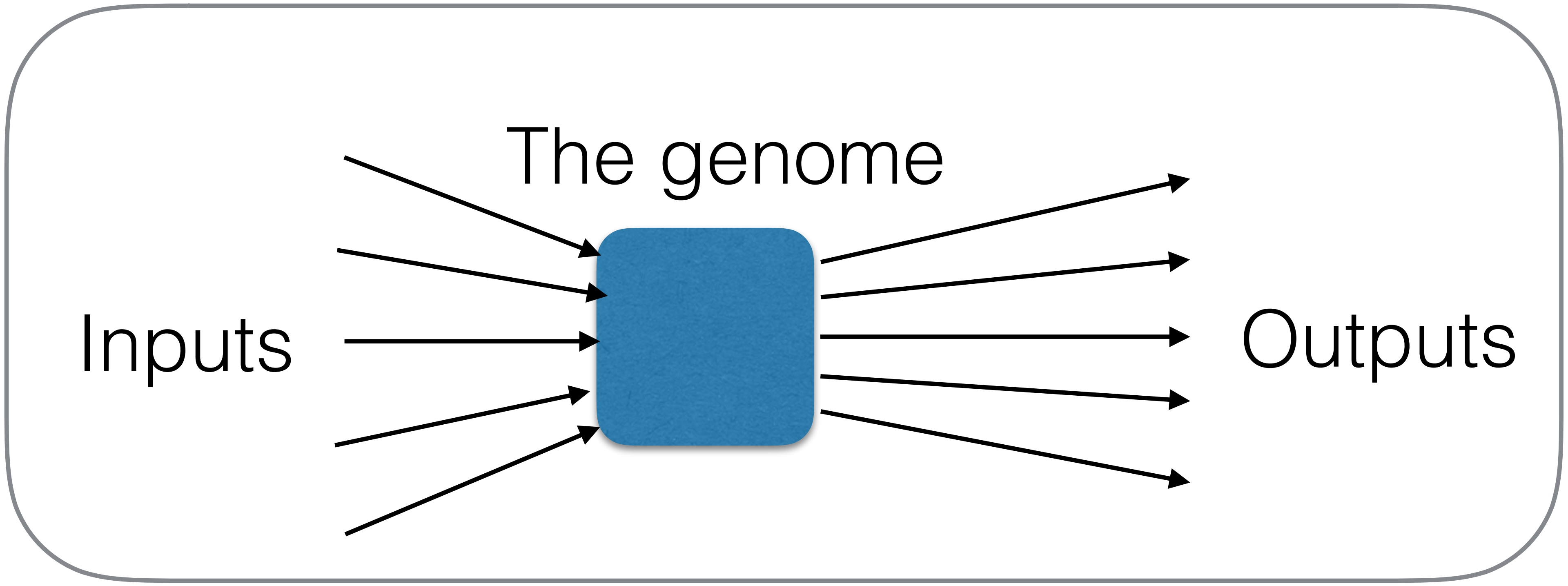
- The input proteins modify the state of the genome to produce output proteins.
- The relationship between these two is extremely complex.
- It is intimately involved with neural development.



A cell

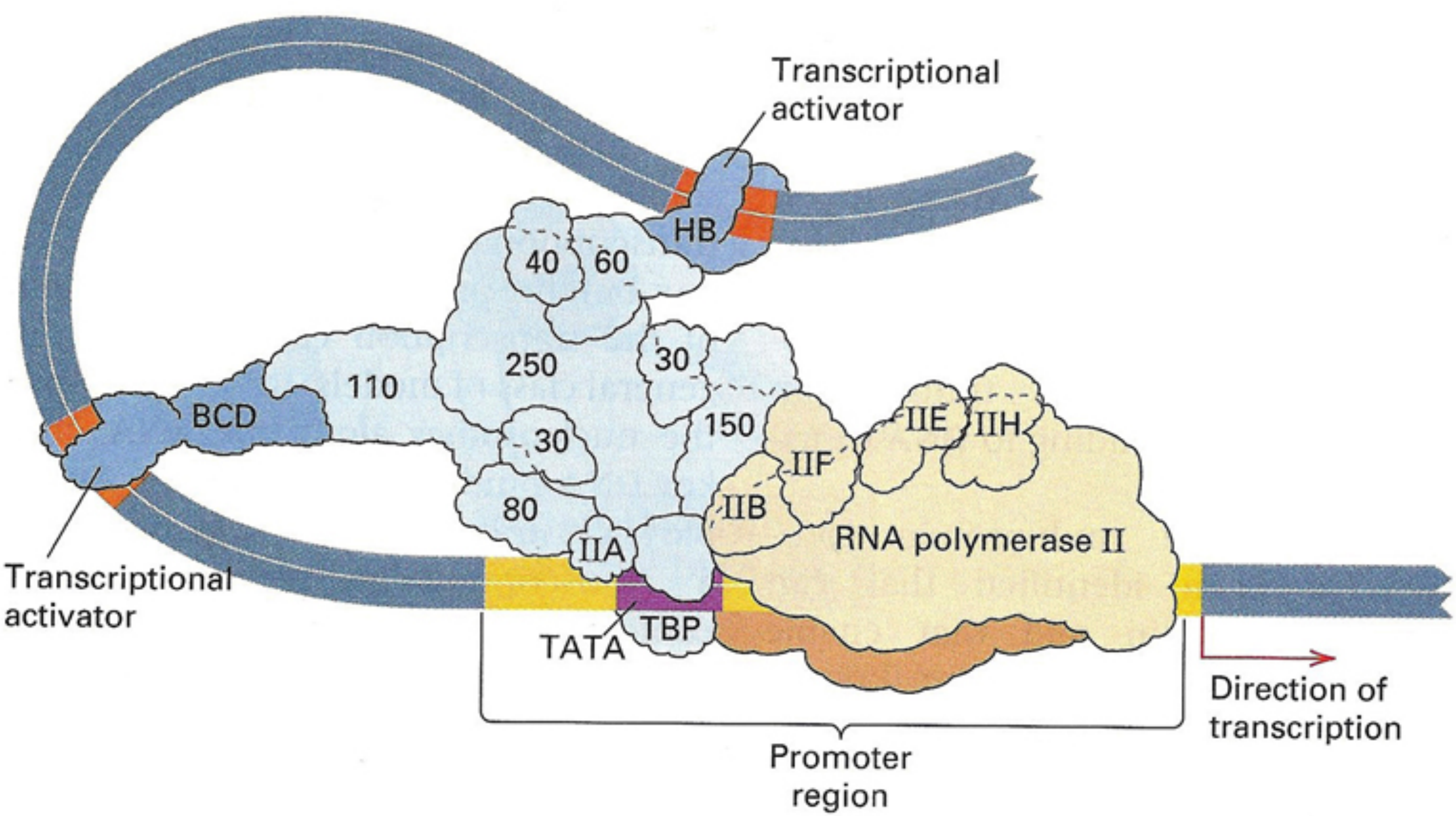
Gene regulation

- The input proteins modify the state of the genome to produce output proteins.
- The relationship between these two is extremely complex.
- It is intimately involved with neural development.



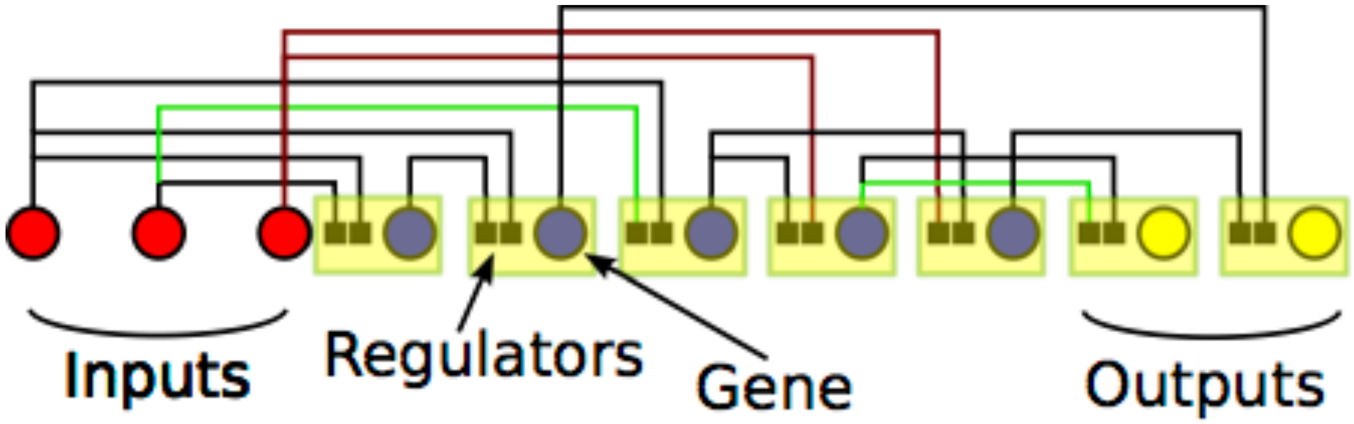
A cell

Activation of eukaryotic transcription



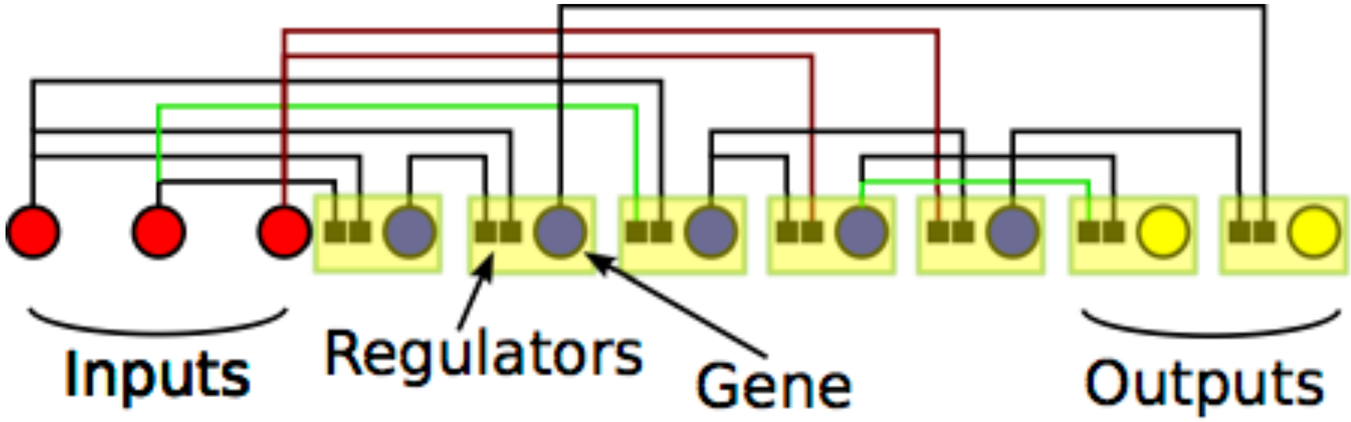
Biological Science, 4th edition, Freeman, 2010

Gene networks



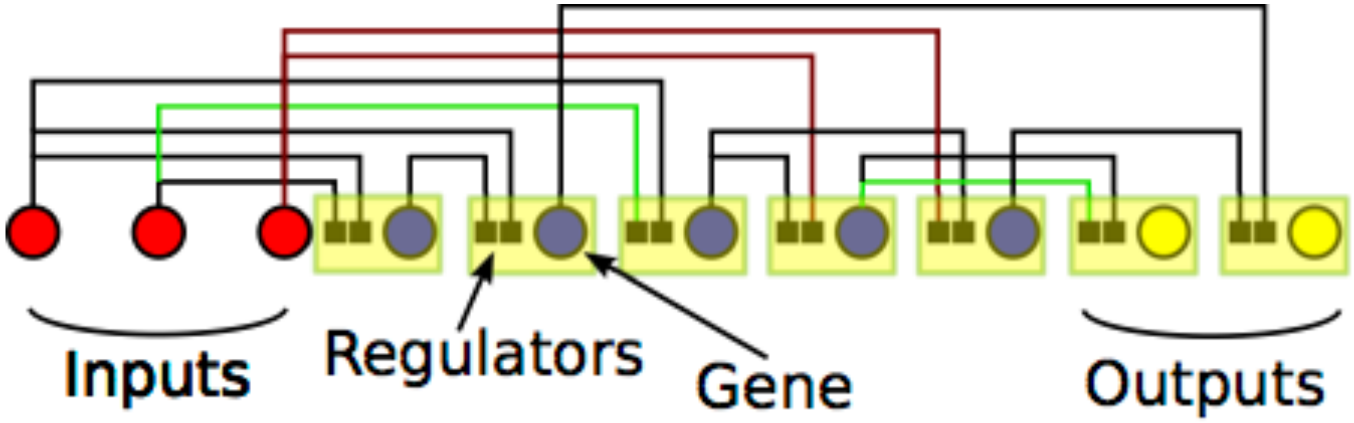
Gene networks

- Gene expression is controlled by a highly complex mechanism, including other genes.

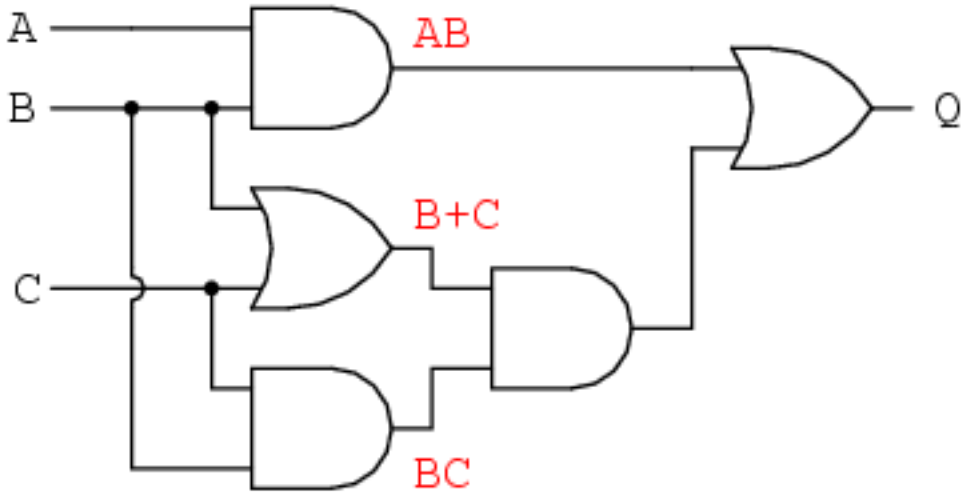


Gene networks

- Gene expression is controlled by a highly complex mechanism, including other genes.
- Some regulatory proteins enhance and others repress transcription, by binding to regulatory elements.



Analogous to digital logic



Relatively small number of inputs and outputs

Complication:
“Junk DNA”

Complication: “Junk DNA”

Most mammalian genomic transcripts do not directly code for proteins and only approximately 5% of the bases can be confidently identified as being under evolutionary constraint.

Complication: “Junk DNA”

Most mammalian genomic transcripts do not directly code for proteins and only approximately 5% of the bases can be confidently identified as being under evolutionary constraint.

The other 95% has been thought of as just due the random nature of evolution. The cost for it being there is not large.

Complication: "Junk DNA"

Most mammalian genomic transcripts do not directly code for proteins and only approximately 5% of the bases can be confidently identified as being under evolutionary constraint.

The other 95% has been thought of as just due the random nature of evolution. The cost for it being there is not large.

The salamander as 10 times our DNA

A pine tree has 7 times

Yet the bladder wort has only 3% noncoding DNA

Complication: “Junk DNA”

Most mammalian genomic transcripts do not directly code for proteins and only approximately 5% of the bases can be confidently identified as being under evolutionary constraint.

The other 95% has been thought of as just due the random nature of evolution. The cost for it being there is not large.

The salamander as 10 times our DNA

A pine tree has 7 times

Yet the bladder wort has only 3% noncoding DNA

ENCODE project

Encyclopedia of DNA Elements

ENCODE project

Encyclopedia of DNA Elements

- Concluded 80% of the genome has biochemical activity.

ENCODE project

Encyclopedia of DNA Elements

- Concluded 80% of the genome has biochemical activity.
- It is pervasively transcribed.

ENCODE project

Encyclopedia of DNA Elements

- Concluded 80% of the genome has biochemical activity.
- It is pervasively transcribed.
- However it is largely not evolutionarily conserved.

ENCODE project

Encyclopedia of DNA Elements

- Concluded 80% of the genome has biochemical activity.
- It is pervasively transcribed.
- However it is largely not evolutionarily conserved.
- Is most of doing anything “useful”?

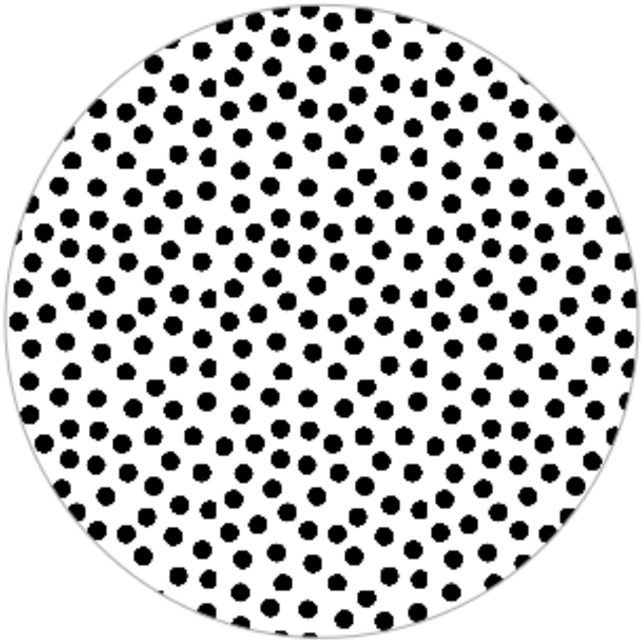
ENCODE project

Encyclopedia of DNA Elements

- Concluded 80% of the genome has biochemical activity.
- It is pervasively transcribed.
- However it is largely not evolutionarily conserved.
- Is most of doing anything “useful”?
- There appear to be quite a few examples, but most of the function of it is still unknown.

Does it “get in the way”?

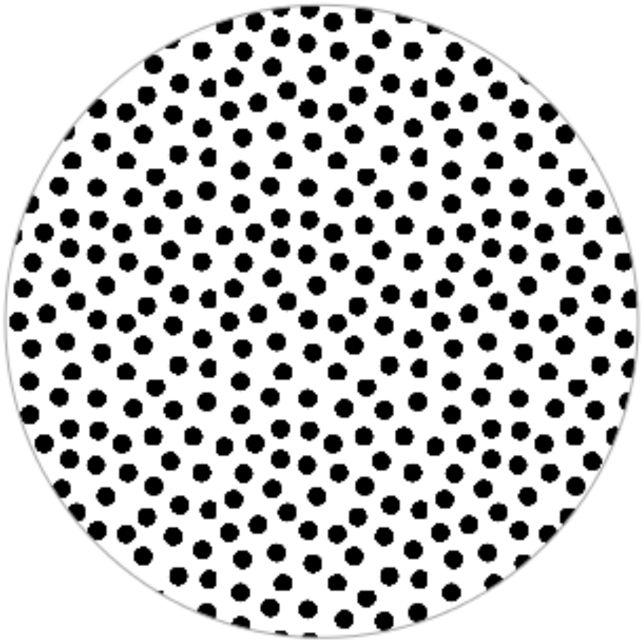
Back of the envelope



Does it “get in the way”?

Back of the envelope

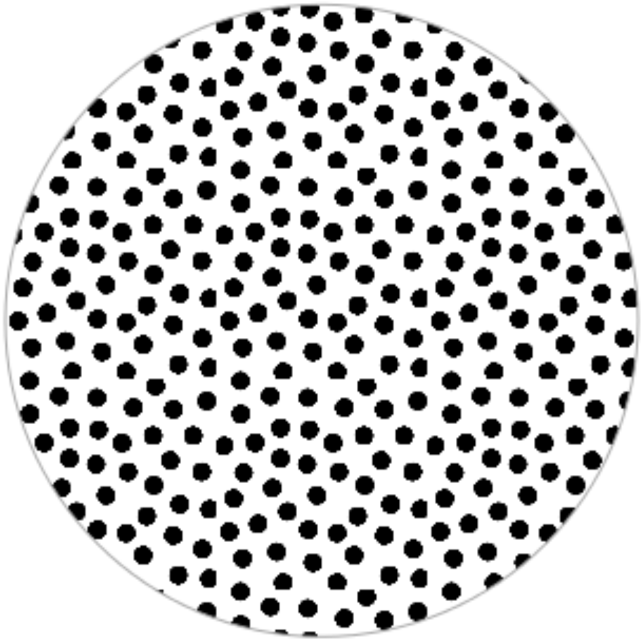
The number of mRNA in human cells ~ 500,000



Does it “get in the way”?

Back of the envelope

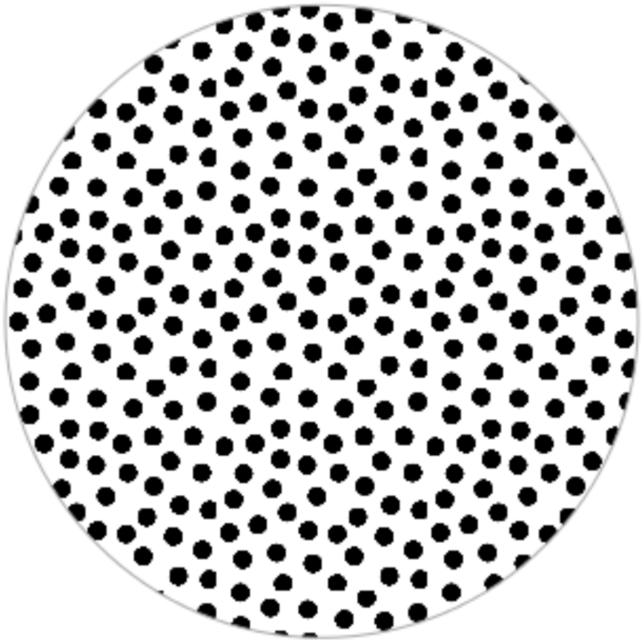
The number of mRNA in human cells ~ 500,000
Amount of non-coding RNA is ~1/10 of that



Does it “get in the way”?

Back of the envelope

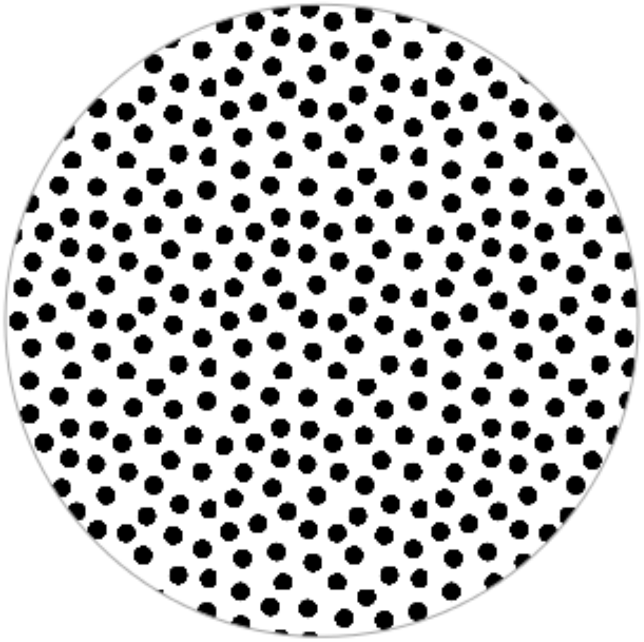
The number of mRNA in human cells ~ 500,000
Amount of non-coding RNA is ~1/10 of that
1/2 localized to nucleus, radius = 3 µm



Does it “get in the way”?

Back of the envelope

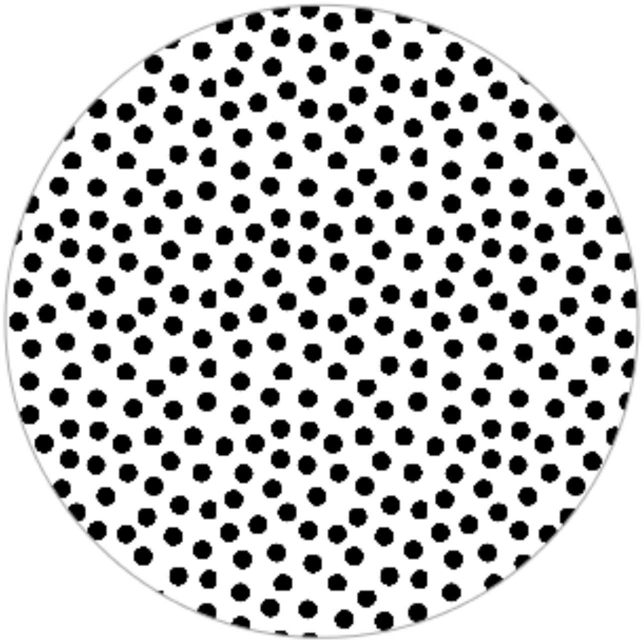
The number of mRNA in human cells ~ 500,000
Amount of non-coding RNA is ~1/10 of that
1/2 localized to nucleus, radius = 3 µm
=> about 20,000 ncRNA/nucleus



Does it “get in the way”?

Back of the envelope

The number of mRNA in human cells ~ 500,000
Amount of non-coding RNA is ~1/10 of that
1/2 localized to nucleus, radius = 3 µm
=> about 20,000 ncRNA/nucleus

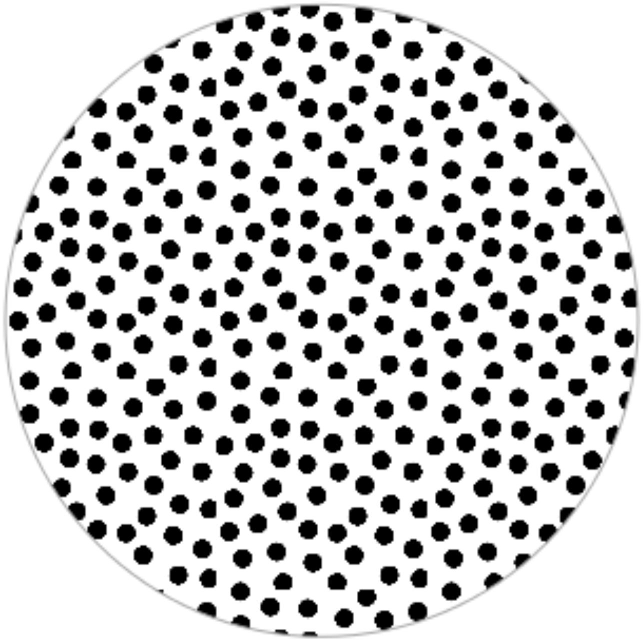


Does it “get in the way”?

Back of the envelope

The number of mRNA in human cells ~ 500,000
Amount of non-coding RNA is ~1/10 of that
1/2 localized to nucleus, radius = 3 µm
=> about 20,000 ncRNA/nucleus

=> distance of 0.18 micron between them

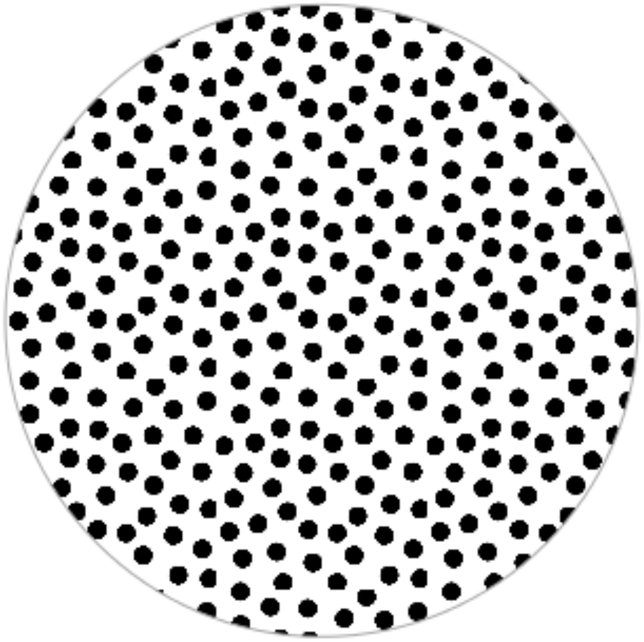


Does it “get in the way”?

Back of the envelope

The number of mRNA in human cells ~ 500,000
Amount of non-coding RNA is ~1/10 of that
1/2 localized to nucleus, radius = 3 µm
=> about 20,000 ncRNA/nucleus

=> distance of 0.18 micron between them

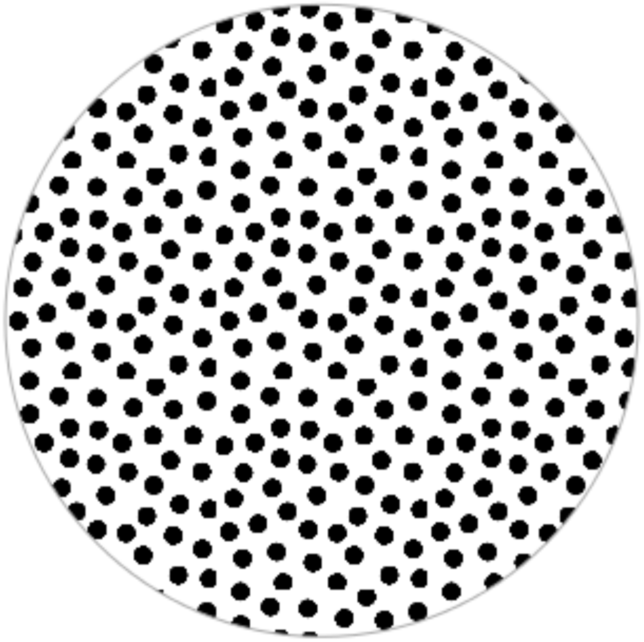


Does it “get in the way”?

Back of the envelope

The number of mRNA in human cells ~ 500,000
Amount of non-coding RNA is ~1/10 of that
1/2 localized to nucleus, radius = 3 µm
=> about 20,000 ncRNA/nucleus

=> distance of 0.18 micron between them



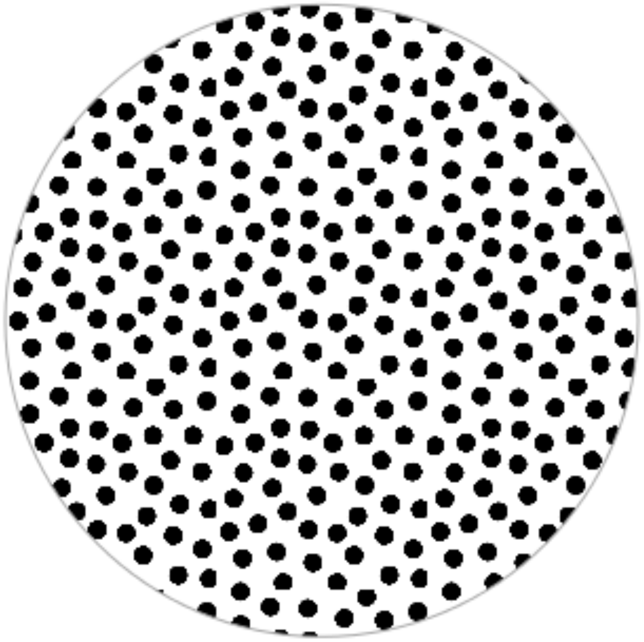
D = 0.1 $\mu m^2/s$

Does it “get in the way”?

Back of the envelope

The number of mRNA in human cells ~ 500,000
Amount of non-coding RNA is ~1/10 of that
1/2 localized to nucleus, radius = 3 µm
=> about 20,000 ncRNA/nucleus

=> distance of 0.18 micron between them



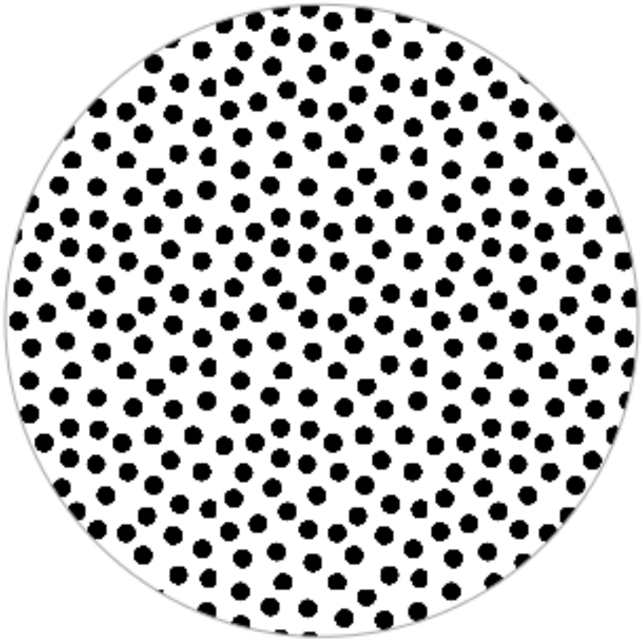
D = 0.1 $\mu m^2/s$

Does it “get in the way”?

Back of the envelope

The number of mRNA in human cells ~ 500,000
Amount of non-coding RNA is ~1/10 of that
1/2 localized to nucleus, radius = 3 µm
=> about 20,000 ncRNA/nucleus

=> distance of 0.18 micron between them



$D = 0.1\ \mu m^2/s$

Size of ncRNA ~ 40 nm

Does it “get in the way”?

Back of the envelope

The number of mRNA in human cells ~ 500,000

Amount of non-coding RNA is ~1/10 of that

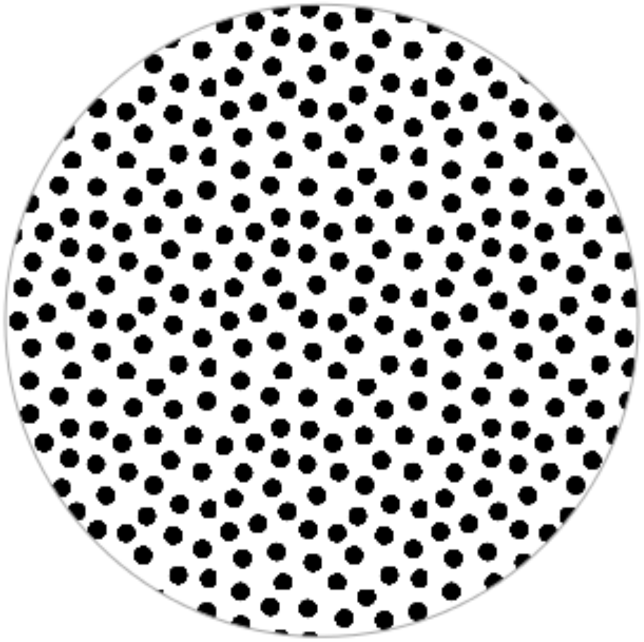
1/2 localized to nucleus, radius = 3 µm

=> about 20,000 ncRNA/nucleus

=> distance of 0.18 micron between them

D = 0.1 $\mu m^2/s$

Size of ncRNA ~ 40 nm

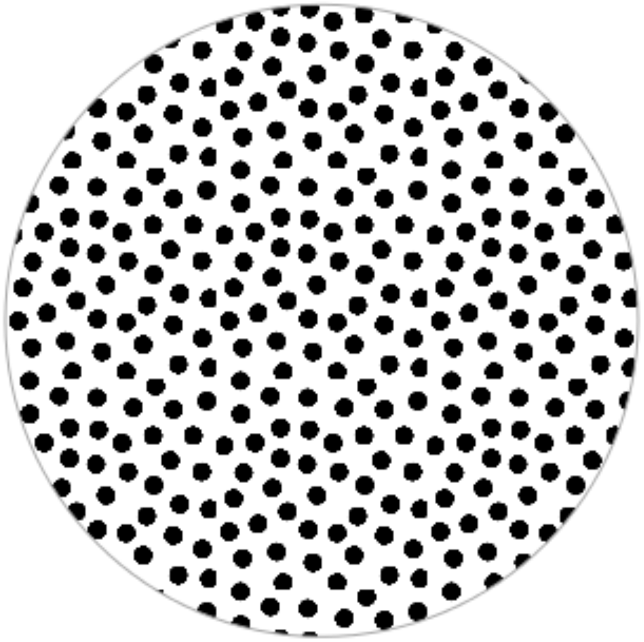


Does it “get in the way”?

Back of the envelope

The number of mRNA in human cells ~ 500,000
Amount of non-coding RNA is ~1/10 of that
1/2 localized to nucleus, radius = 3 µm
=> about 20,000 ncRNA/nucleus

=> distance of 0.18 micron between them



D = 0.1 µm²/s

Size of ncRNA ~ 40 nm

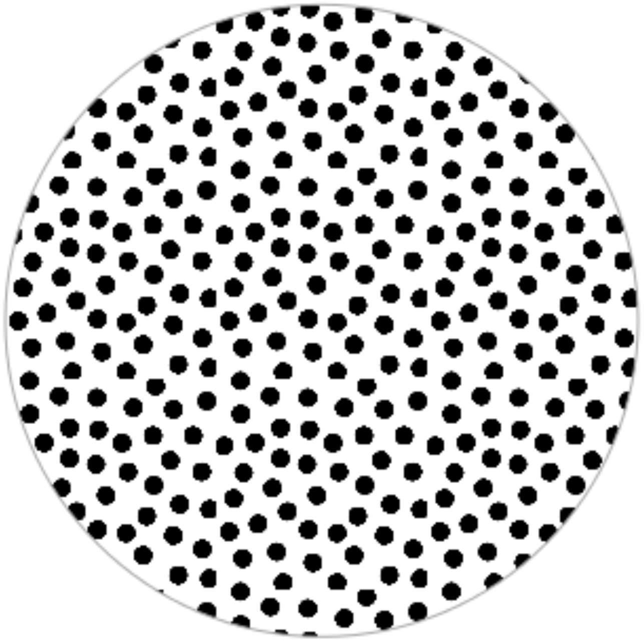
=> Timescale for interaction ~ 0.25 s

Does it "get in the way"?

Back of the envelope

The number of mRNA in human cells ~ 500,000
Amount of non-coding RNA is ~1/10 of that
1/2 localized to nucleus, radius = 3 µm
=> about 20,000 ncRNA/nucleus

=> distance of 0.18 micron between them

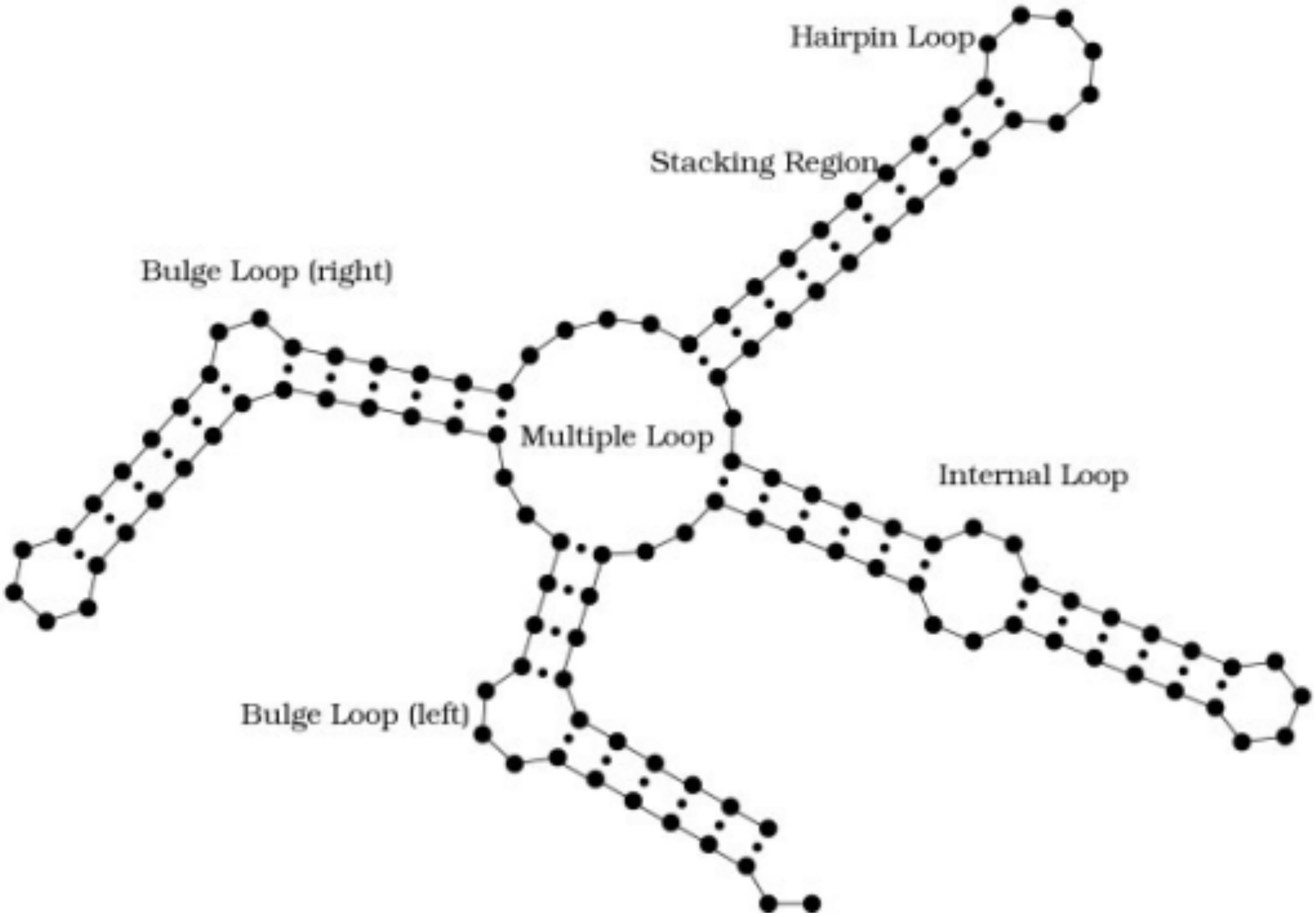


D = 0.1 $\mu m^2/s$

Size of ncRNA ~ 40 nm

=> Timescale for interaction ~ 0.25 s

RNA is “sticky”. e.g. It has a lot of secondary structure



Non-coding RNA takes about 30 - 60 minutes to degrade!

Non-coding RNA takes about 30 - 60 minutes to degrade!

It’s certainly interacting with itself and lots of other stuff in the nucleus!

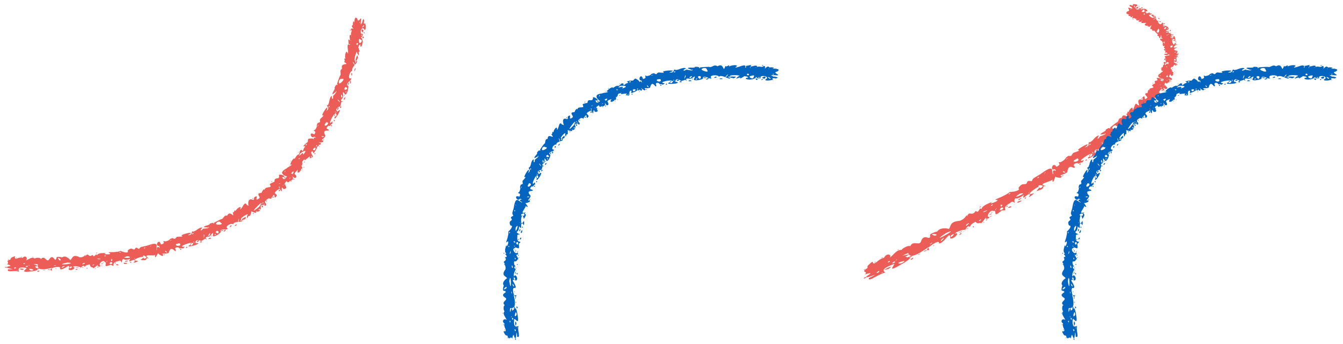
Non-coding RNA takes about 30 - 60 minutes to degrade!

It’s certainly interacting with itself and lots of other stuff in the nucleus!

On physical grounds, it is doing something!

RNA Chemical equilibration

$$i + j \rightleftharpoons ij$$



Total concentration

$$\rho_i = \frac{C_i}{1 + \sum_j \rho_j K_{ij}}$$

$$K_{ij} = \rho_{ij} / \rho_i \rho_j$$

unbound concentration

Equilibrium constant

Relaxation

$$\tau_\rho \frac{d\rho_i}{dt} = -\rho_i + \frac{C_i}{1 + \sum_j \rho_j K_{ij}}$$

Equilibration time, of order 1s

RNA creation

RNA is both degraded and created. For species i

$$\tau_C \frac{dC_i}{dt} = -C_i + f(C_i, \{\rho_k\})$$

RNA creation

RNA is both degraded and created. For species i

$$\tau_C \frac{dC_i}{dt} = -C_i + f(C_i, \{\rho_k\})$$

Time scale is of order 30 to 60 minutes

RNA creation

RNA is both degraded and created. For species i

$$\tau_C \frac{dC_i}{dt} = -C_i + f(C_i, \{\rho_k\})$$

Time scale is of order 30 to 60 minutes

Creation rate, f, is regulated by the total concentration of i and the concentration of unbound molecules.

Assumption:

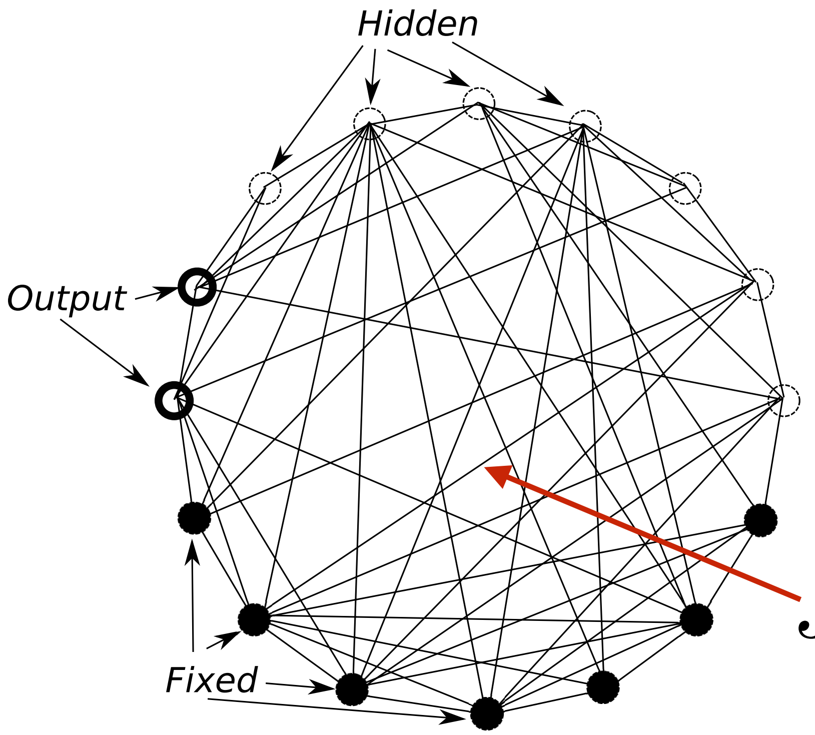
All this “junk” will evolve to do something useful

Assumption:

All this “junk” will evolve to do something useful

i.e. Inputs to the genome will give sensible outputs

Boltzmann Machine & Hopfield Model

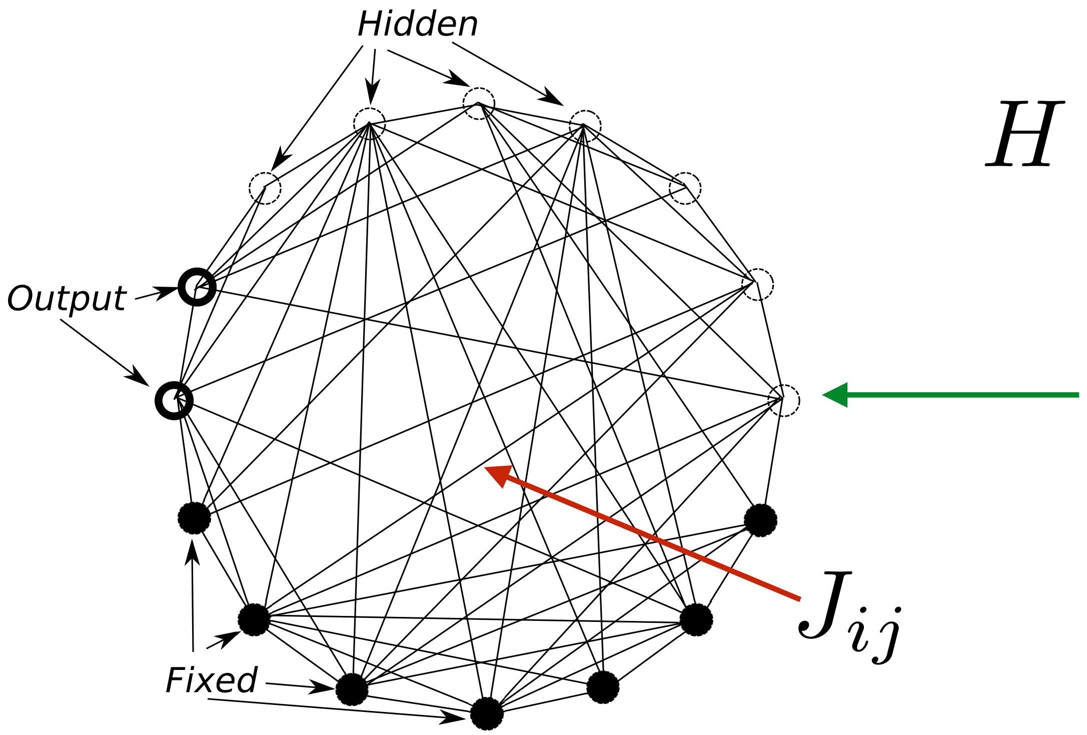


$$H = - \sum_{i=1,j=1}^{N} J_{ij} s_i s_j$$

$$s_i = \pm 1$$

J_{ij}

Boltzmann Machine & Hopfield Model



$$H = -\sum_{i=1,j=1}^{N} J_{ij} s_i s_j$$

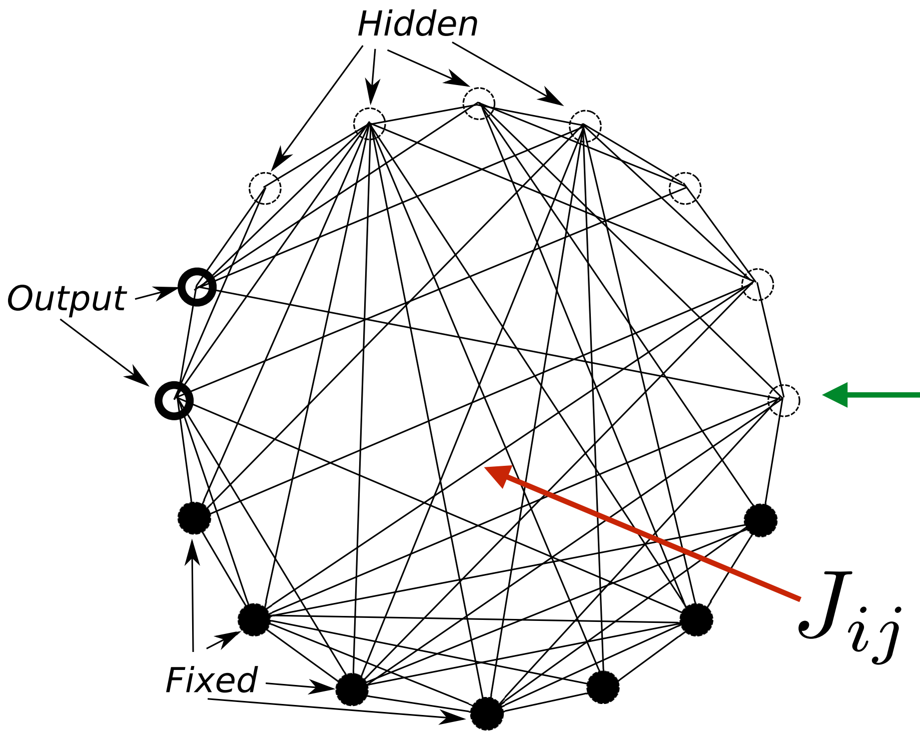
$$s_i = \pm 1$$

J_{ij}

$J_{ij}'\mathrm{s}$

Boltzmann Machine & Hopfield Model

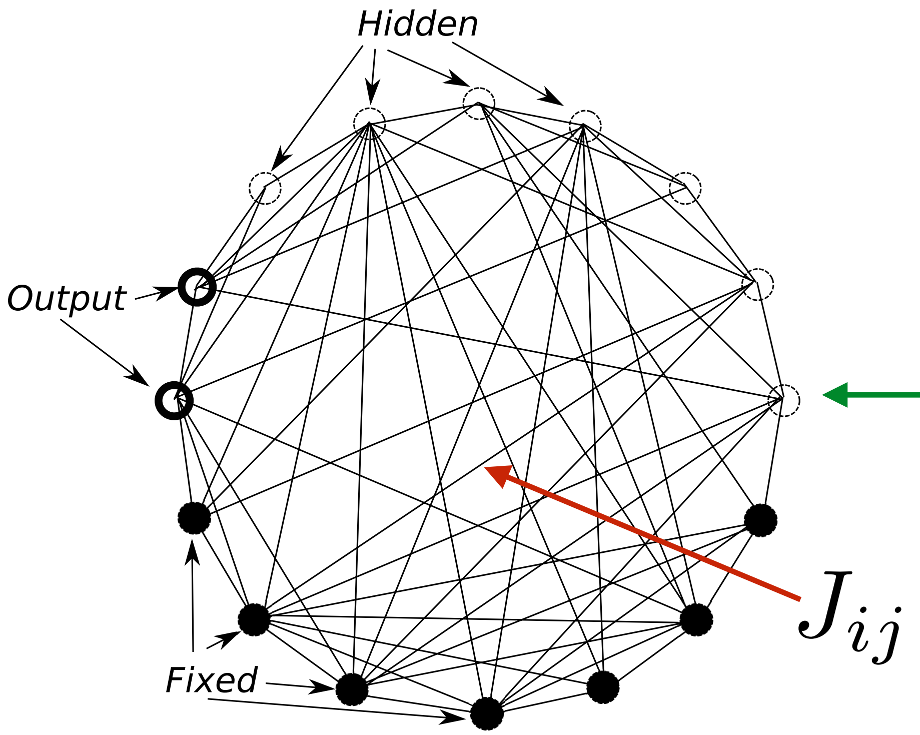
$$H = - \sum_{i=1,j=1}^{N} J_{ij} s_i s_j$$



$$s_i = \pm 1$$

J_{ij}'s chosen to give optimal outputs given inputs

Boltzmann Machine & Hopfield Model

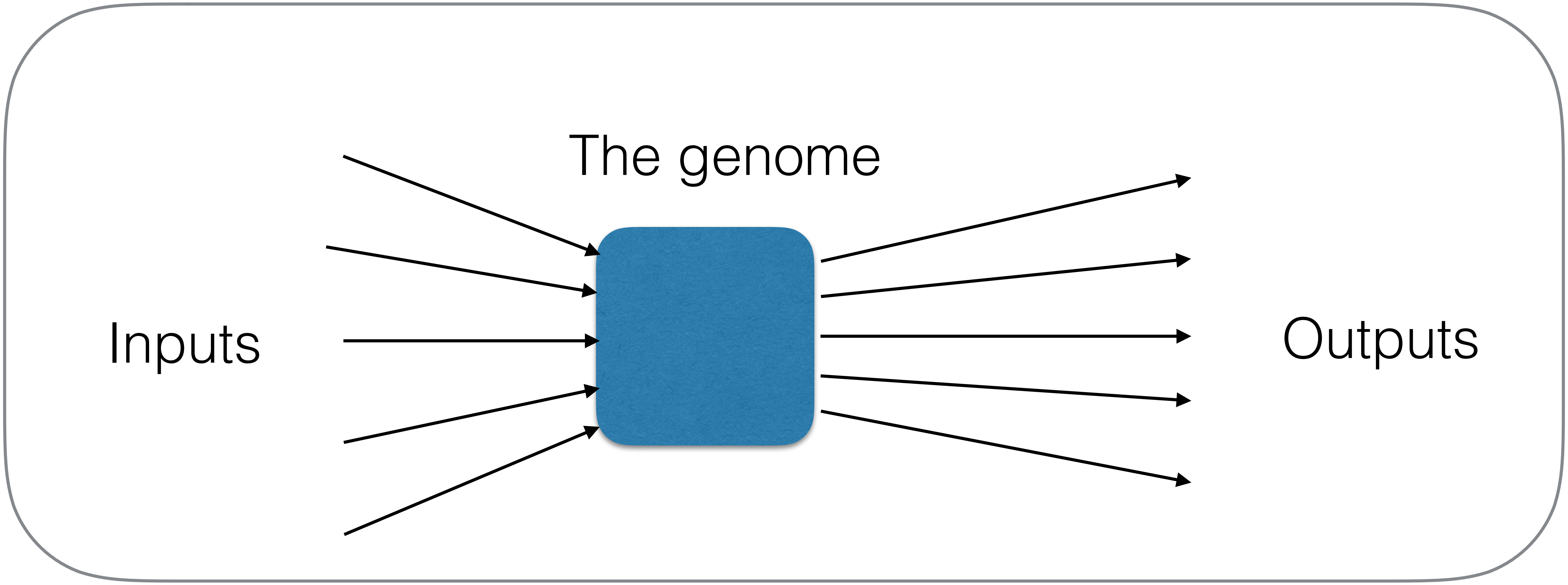


$$H = -\sum_{i=1,j=1}^{N} J_{ij} s_i s_j$$

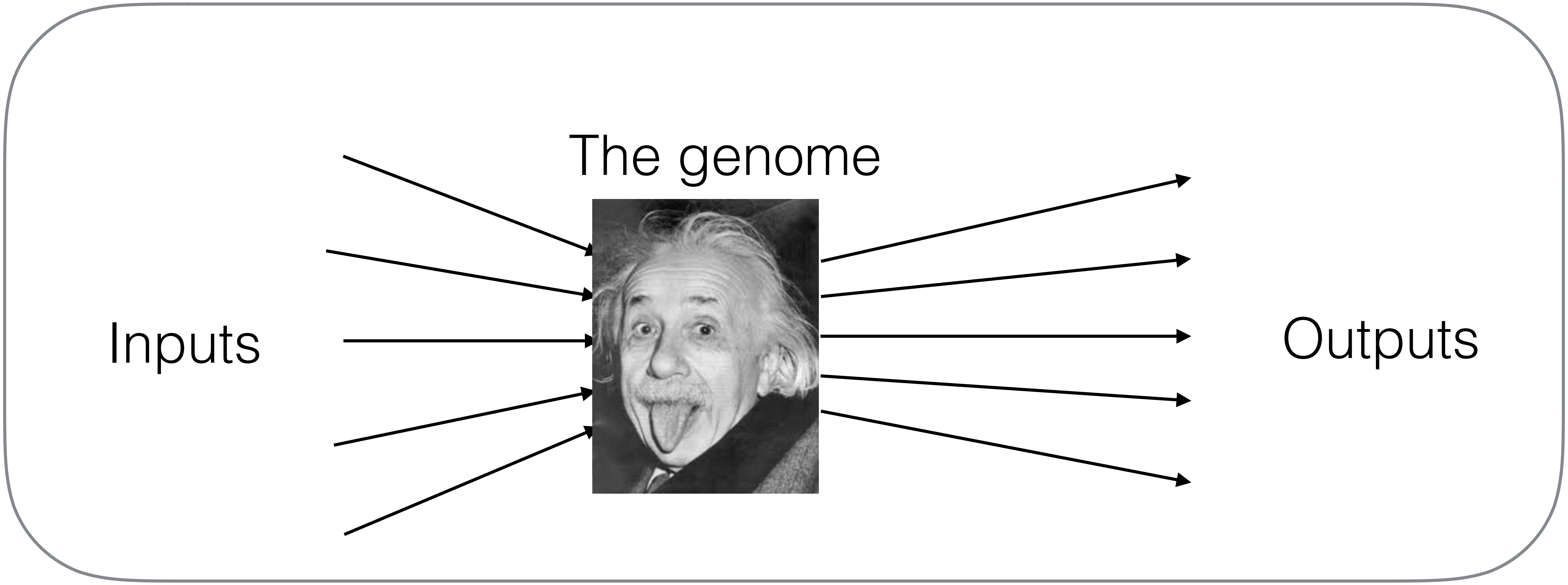
$$s_i = \pm 1$$

$J_{ij}{}'\mathrm{s}$ chosen to give optimal outputs given inputs

$$s_i = \tanh(\frac{\beta}{N}\sum_j J_{ij} s_j)$$



A cell



The genome

Inputs

Outputs

A smart cell

Neural net / RNA mapping

spin $\longrightarrow$ $s_i \Longleftrightarrow \rho_i$ $\longleftarrow$ Unbound concentration

spin coupling $\longrightarrow$ $J_{ij} \Longleftrightarrow K_{ij}$ $\longleftarrow$ Equilibrium Constant

$$\rho_i = \delta \frac{1 + s_i}{2} + b$$

$$K_{ij} = \epsilon \frac{1 + J_{ij}}{2} + a$$

$t \to \infty$: Want RNA’s to give neural net behavior

Choose creation rate

Choose creation rate

$$f(C_i, \{\rho_k\}) = \frac{C_i}{\rho_i} S[\frac{4}{\epsilon}(\frac{C_i}{\rho_i} - 1) - 2(1 + 2\frac{a}{\epsilon}) \sum_{j=1}^{N} \rho_j \frac{2(\delta + 2b)}{\epsilon} \sum_j K_{ij} + (\frac{2a}{\epsilon} + 1)(\delta + 2b)N]$$

Choose creation rate

$$f(C_i, \{\rho_k\}) = \frac{C_i}{\rho_i} S[\frac{4}{\epsilon}(\frac{C_i}{\rho_i} - 1) - 2(1 + 2\frac{a}{\epsilon}) \sum_{j=1}^{N} \rho_j \frac{2(\delta + 2b)}{\epsilon} \sum_j K_{ij} + (\frac{2a}{\epsilon} + 1)(\delta + 2b)N]$$

$$S(x) = \frac{\delta}{2}[1 + \tanh(\beta x / N)] + b$$

Choose creation rate

$$f(C_i, \{\rho_k\}) = \frac{C_i}{\rho_i} S[\frac{4}{\epsilon}(\frac{C_i}{\rho_i} - 1) - 2(1 + 2\frac{a}{\epsilon}) \sum_{j=1}^{N} \rho_j \frac{2(\delta + 2b)}{\epsilon} \sum_j K_{ij} + (\frac{2a}{\epsilon} + 1)(\delta + 2b)N]$$

$$S(x) = \frac{\delta}{2}[1 + \tanh(\beta x/N)] + b$$

$t \rightarrow \infty$ Gives: $$s_i = \tanh(\frac{\beta\delta}{N} \sum_j J_{ij} s_j)$$

Choose creation rate

$$f(C_i, \{\rho_k\}) = \frac{C_i}{\rho_i} S[\frac{4}{\epsilon}(\frac{C_i}{\rho_i} - 1) - 2(1 + 2\frac{a}{\epsilon}) \sum_{j=1}^{N} \rho_j \frac{2(\delta + 2b)}{\epsilon} \sum_j K_{ij} + (\frac{2a}{\epsilon} + 1)(\delta + 2b)N]$$

$$S(x) = \frac{\delta}{2}[1 + \tanh(\beta x/N)] + b$$

$t \to \infty$ Gives: $$s_i = \tanh(\frac{\beta\delta}{N} \sum_j J_{ij} s_j)$$

Minimum spin states for Boltzmann machine/ Hopfield

More universal creation

$$f(C_i, \{\rho_k\}) = \frac{C_i}{\rho_i} S(\frac{4}{\epsilon}(\frac{C_i}{\rho_i} - 1) - 2(1 + 2\frac{a}{\epsilon}) \sum_{j=1}^{N} \rho_j - \frac{2(\delta + 2b)}{\epsilon} \sum_j K_{ij} + (\frac{2a}{\epsilon} + 1)(\delta + 2b)N)$$

An ugly mess!

More universal creation

$$f(C_i, \{\rho_k\}) = \frac{C_i}{\rho_i} S(\frac{4}{\epsilon}(\frac{C_i}{\rho_i} - 1) - 2(1 + 2\frac{a}{\epsilon}) \sum_{j=1}^{N} \rho_j - \frac{2(\delta + 2b)}{\epsilon} \sum_j K_{ij} + (\frac{2a}{\epsilon} + 1)(\delta + 2b)N)$$

An ugly mess!

More universal creation

$$f(C_i, \{\rho_k\}) = \frac{C_i}{\rho_i} S(\frac{4}{\epsilon}(\frac{C_i}{\rho_i} - 1) \underbrace{- 2(1 + 2\frac{a}{\epsilon}) \sum_{j=1}^{N} \rho_j - \frac{2(\delta + 2b)}{\epsilon} \sum_j K_{ij} + (\frac{2a}{\epsilon} + 1)(\delta + 2b)N)}$$

An ugly mess!

$$f(C_i, \{\rho_k\}) = \frac{C_i}{\rho_i} S(\frac{4}{\epsilon}(\frac{C_i}{\rho_i} - 1) - A)$$

$\sum_j \rho_j$'s chosen to be target value

More universal creation

$$f(C_i, \{\rho_k\}) = \frac{C_i}{\rho_i} S(\frac{4}{\epsilon}(\frac{C_i}{\rho_i} - 1) - \underbrace{2(1 + 2\frac{a}{\epsilon})\sum_{j=1}^{N} \rho_j - \frac{2(\delta + 2b)}{\epsilon} \sum_j K_{ij} + (\frac{2a}{\epsilon} + 1)(\delta + 2b)N}_{})$$

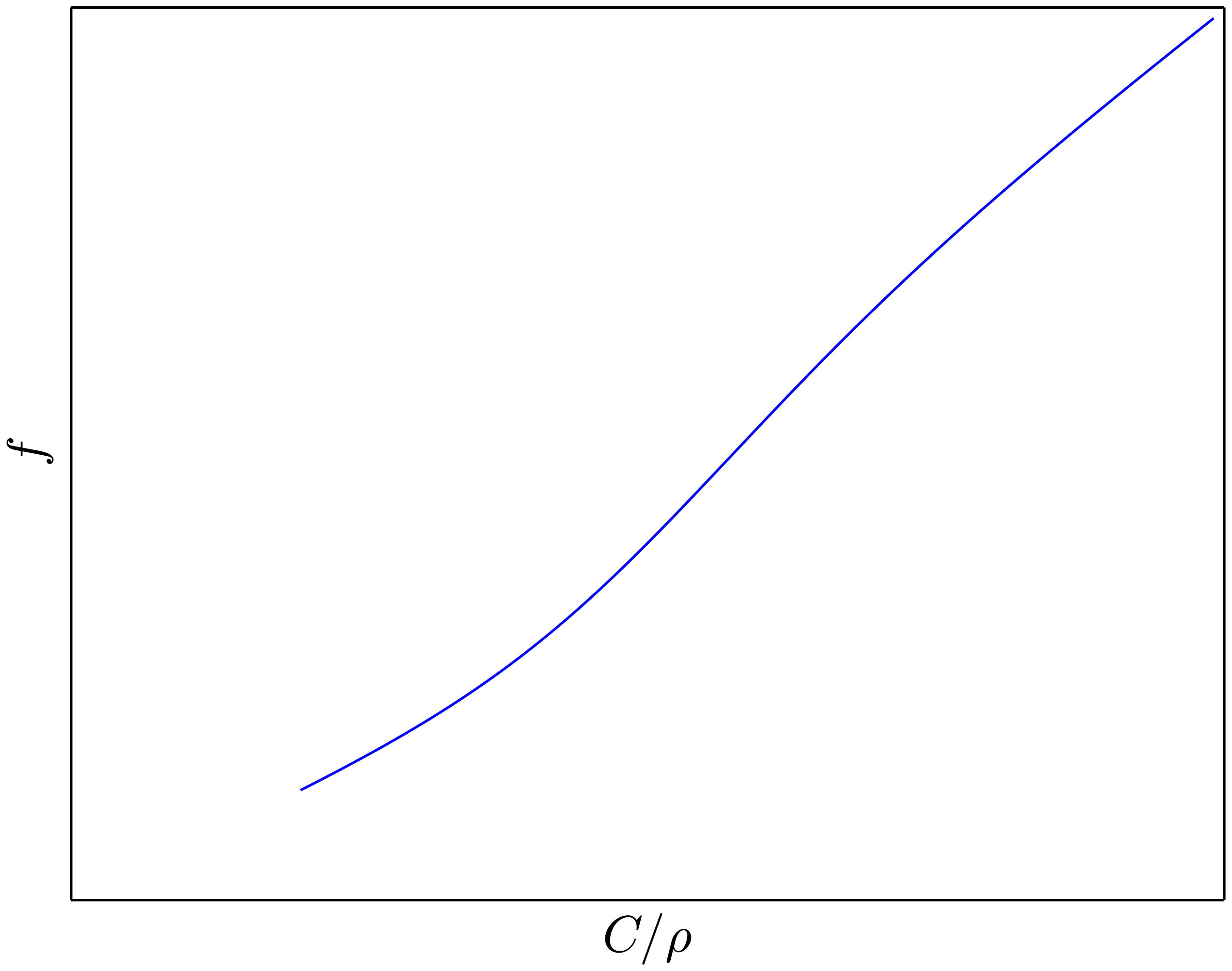
An ugly mess!

$$f(C_i, \{\rho_k\}) = \frac{C_i}{\rho_i} S(\frac{4}{\epsilon}(\frac{C_i}{\rho_i} - 1) - A)$$

$\sum_j \rho_j$'s chosen to be target value

A is a constant

Creation function



Creation rate

Total/unbound RNA

This system gives rise to “Collective Regulation”

This system gives rise to “Collective Regulation”

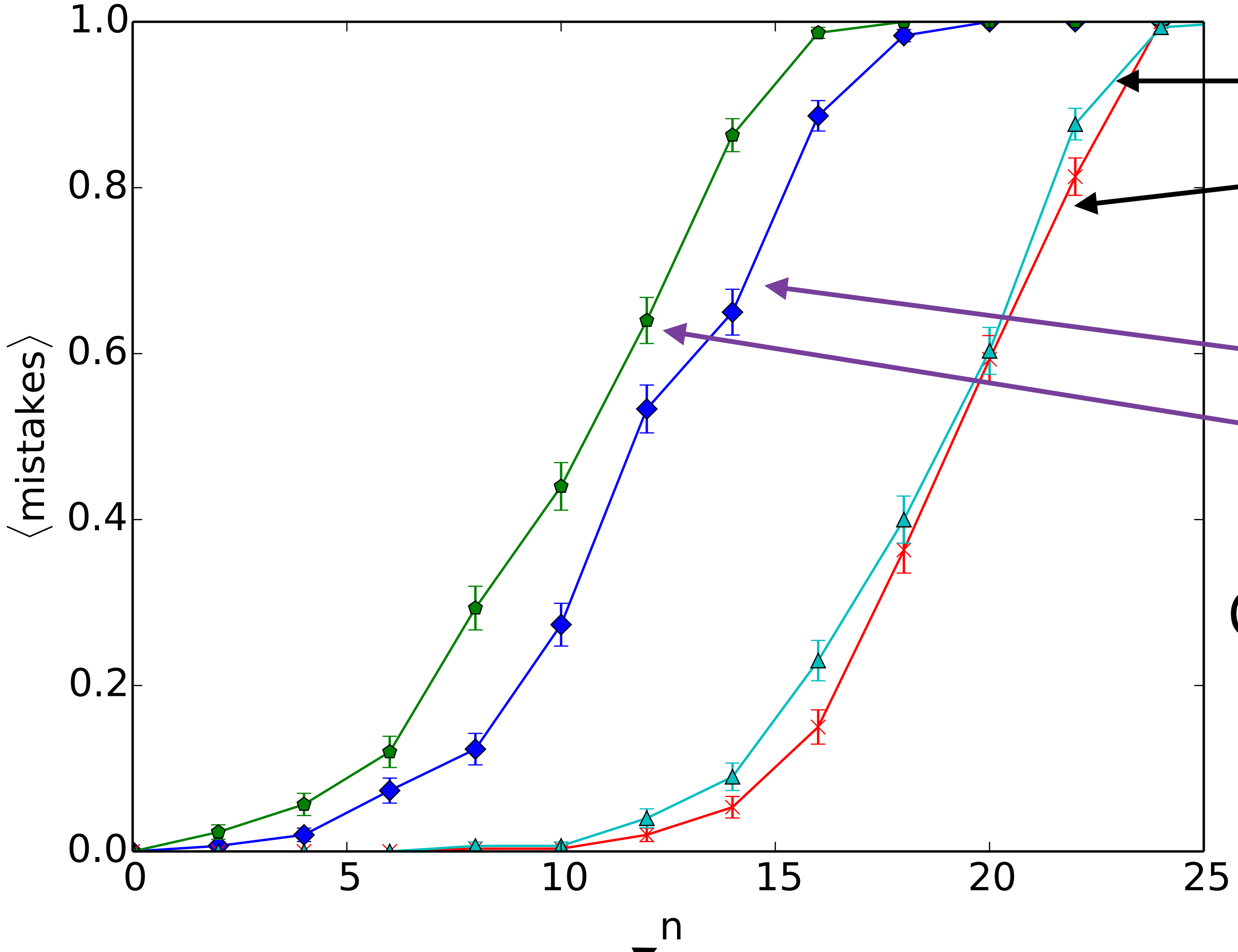
Each interaction has a minute effect but together
the regulate the genome performing precise computations.

This system gives rise to “Collective Regulation”

Each interaction has a minute effect but together the regulate the genome performing precise computations.

Equivalent to a neural network

Tests



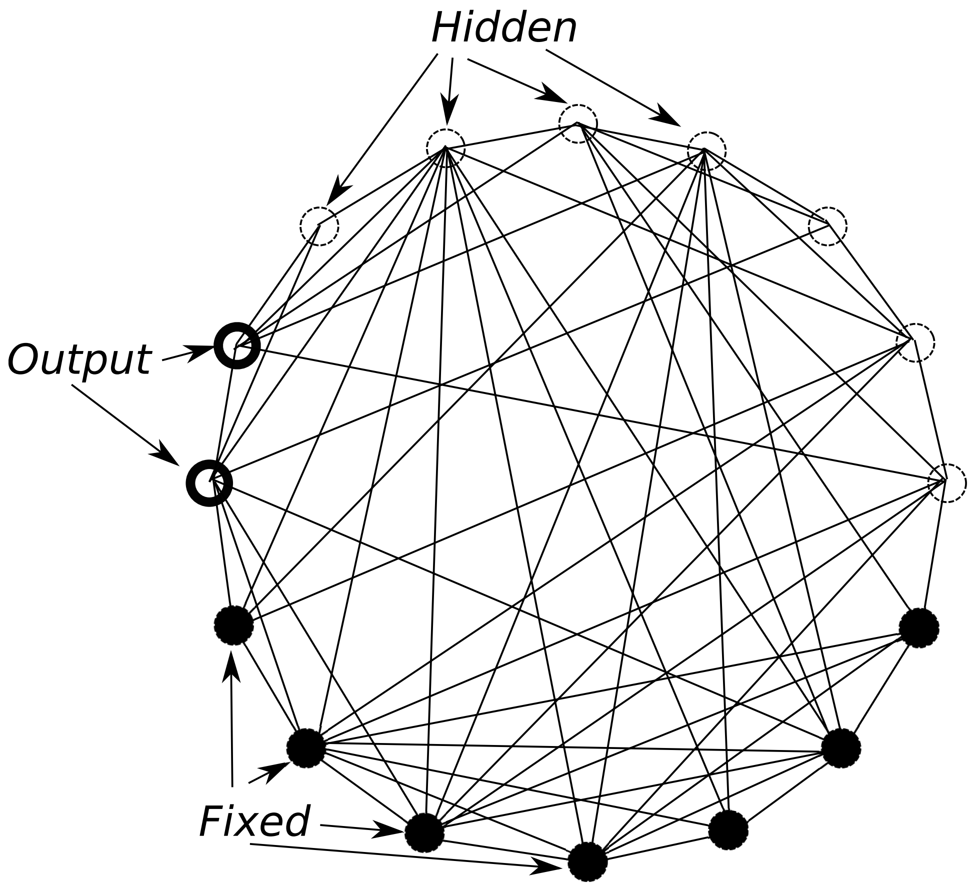
Different $\beta's$

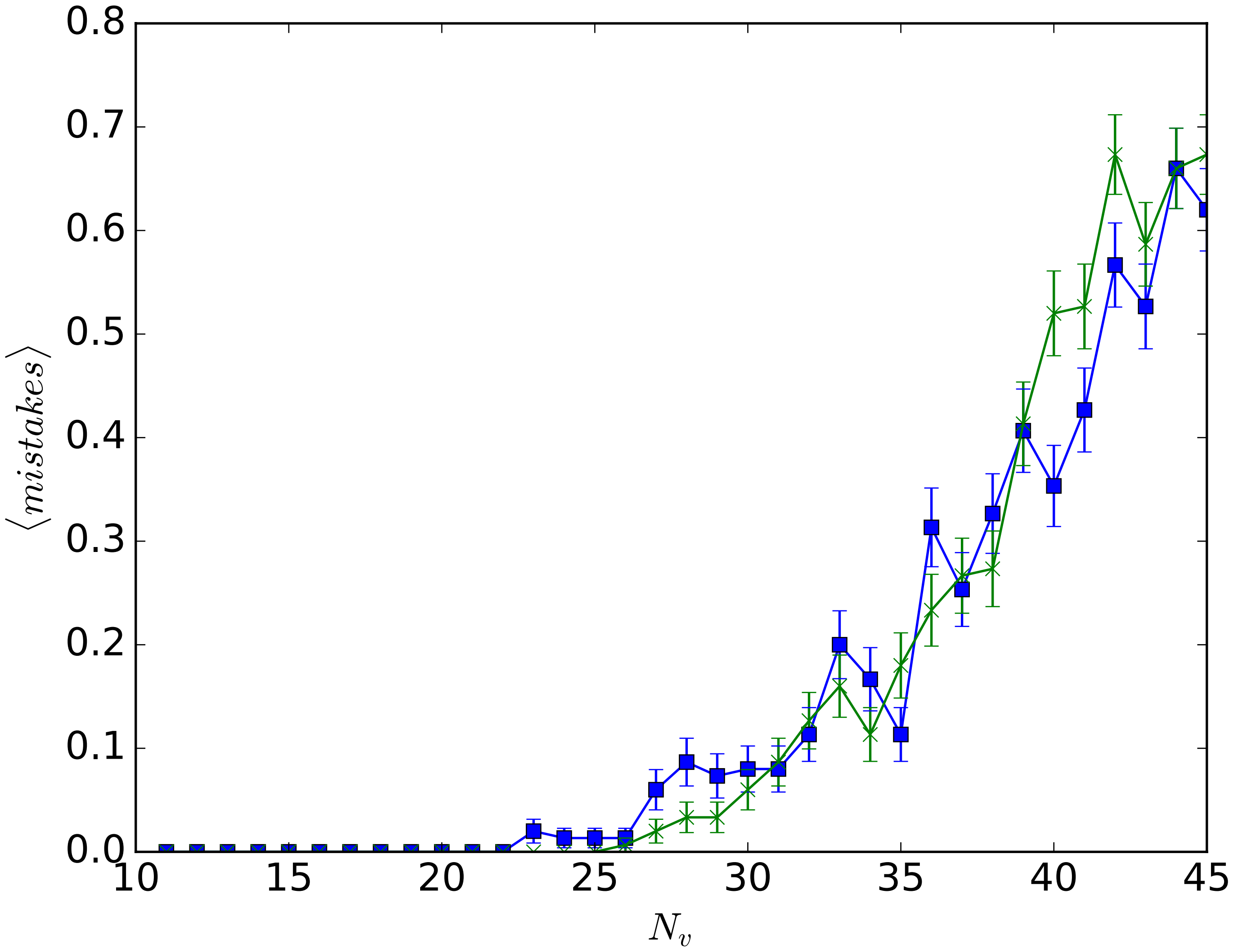
Different initial RNA concentration

Initial Hamming distance

N=50 species

M=3 patterns



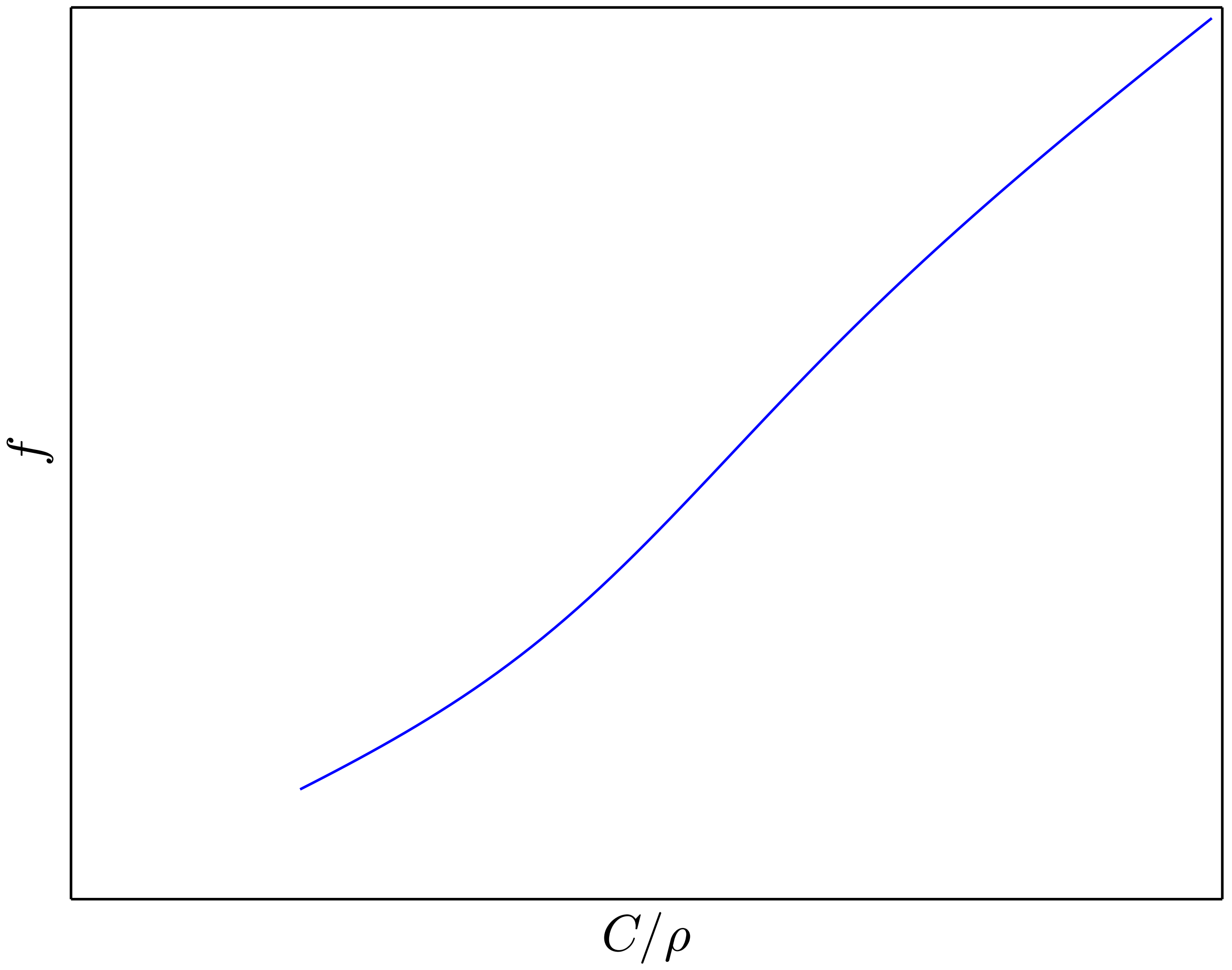


Fraction of mistakes as a function of number of hidden units.

Hidden units are initially scrambled.

50 neurons, 3 patterns. 2 output units.

Is this plausible?



Creation rate

Total/unbound noncoding RNA

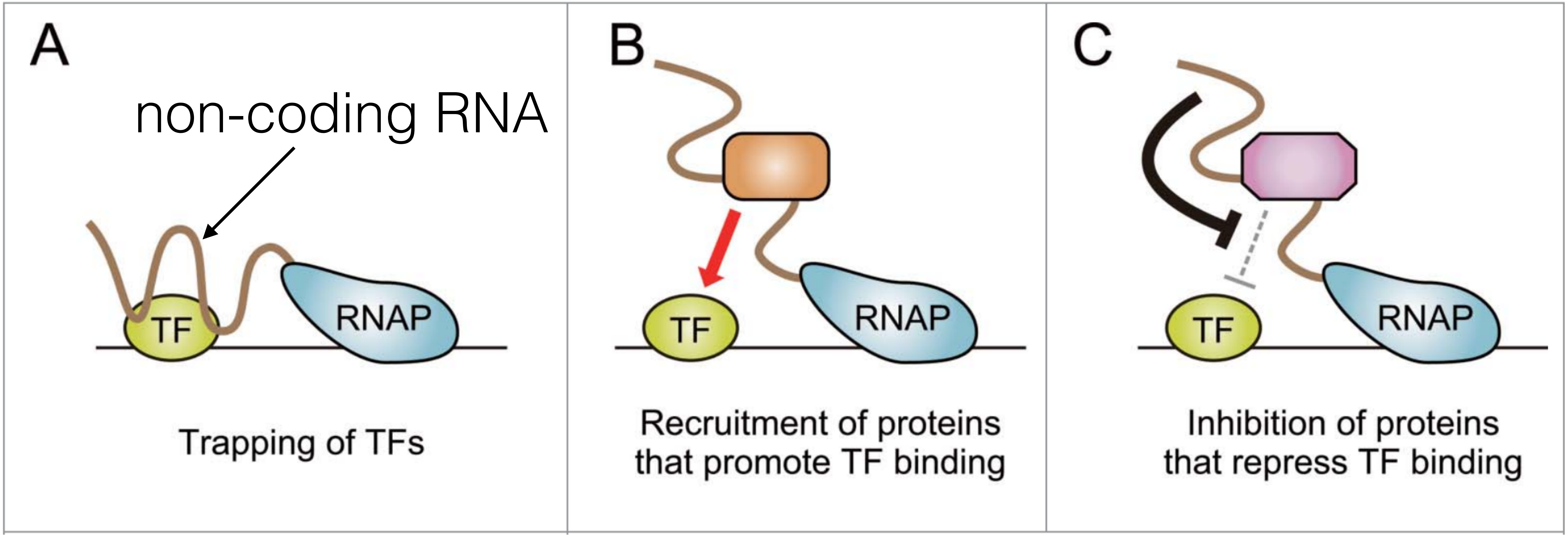
Use a known mechanism

Presence of RNA increases its own transcription

Use a known mechanism

Presence of RNA increases its own transcription

Possible models for how this happens:



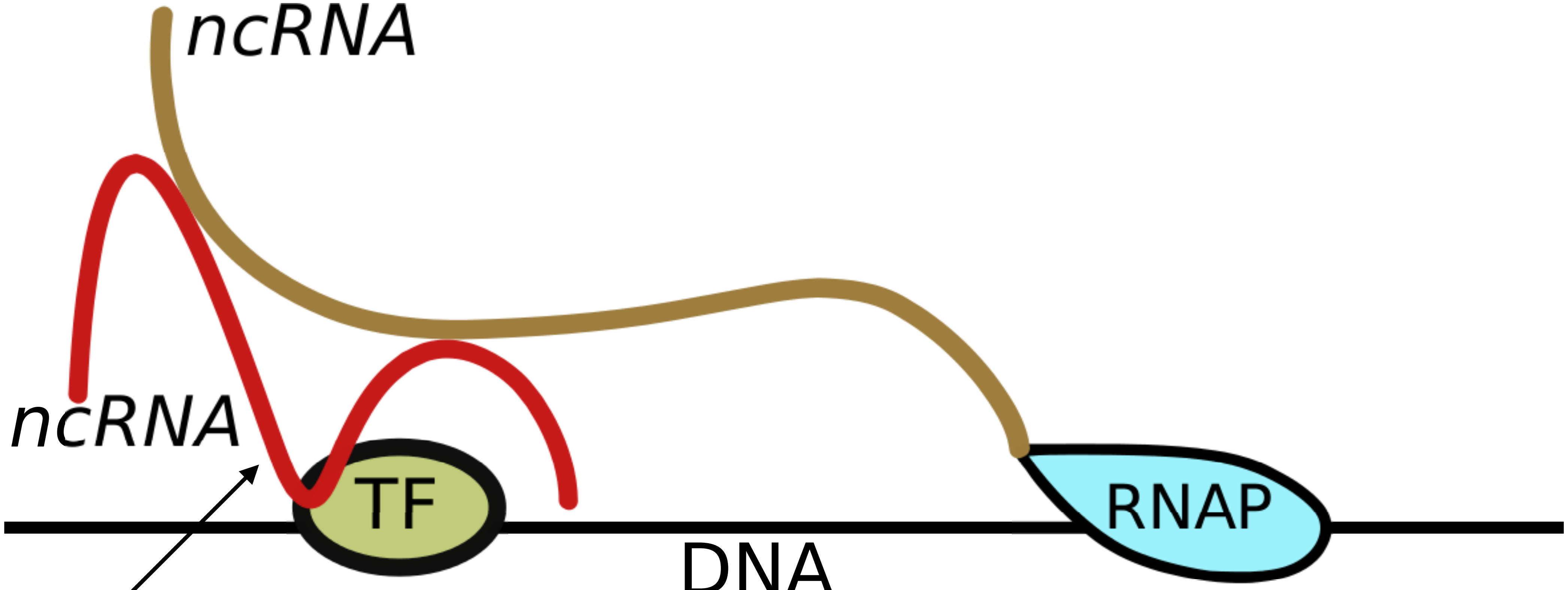
Trapping of TFs

Recruitment of proteins that promote TF binding

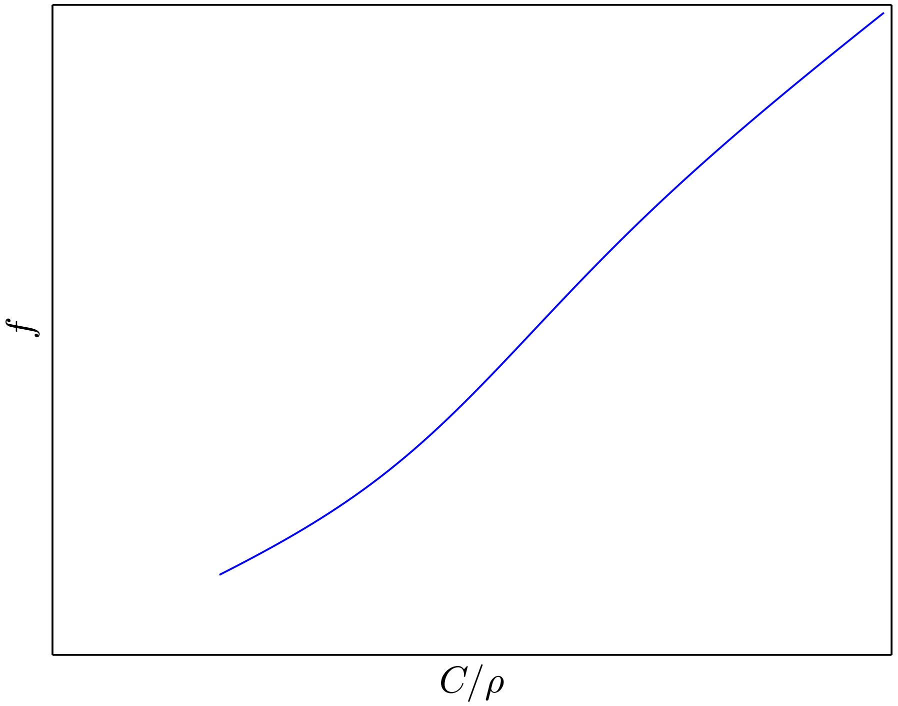
Inhibition of proteins that repress TF binding

“Role of non-coding RNA transcription around gene regulatory elements in transcription factor recruitment”
Naomichi Takemata and Kunihiro Ohta RNA BIOLOGY 2017

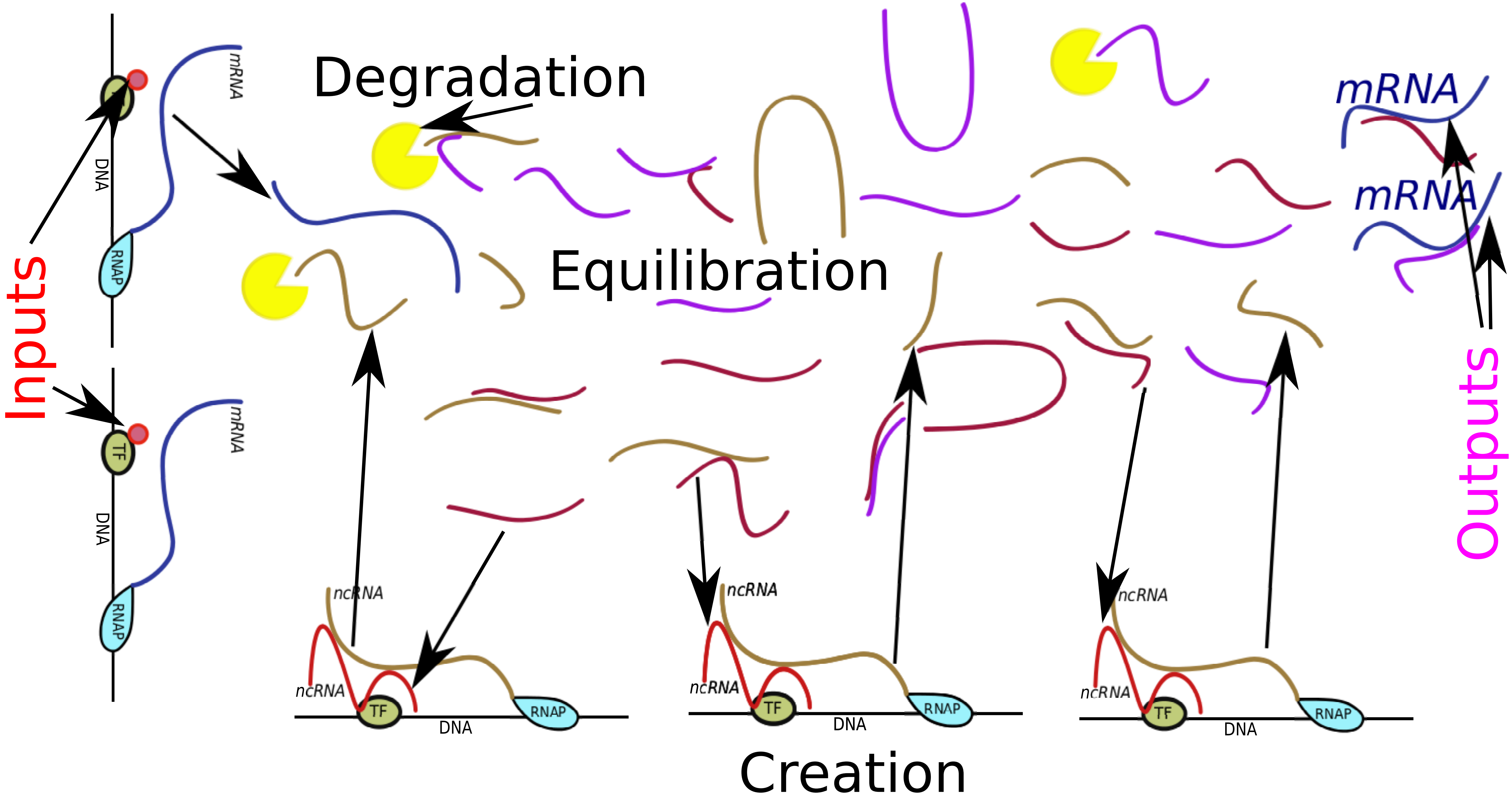
What happens with bound RNA?



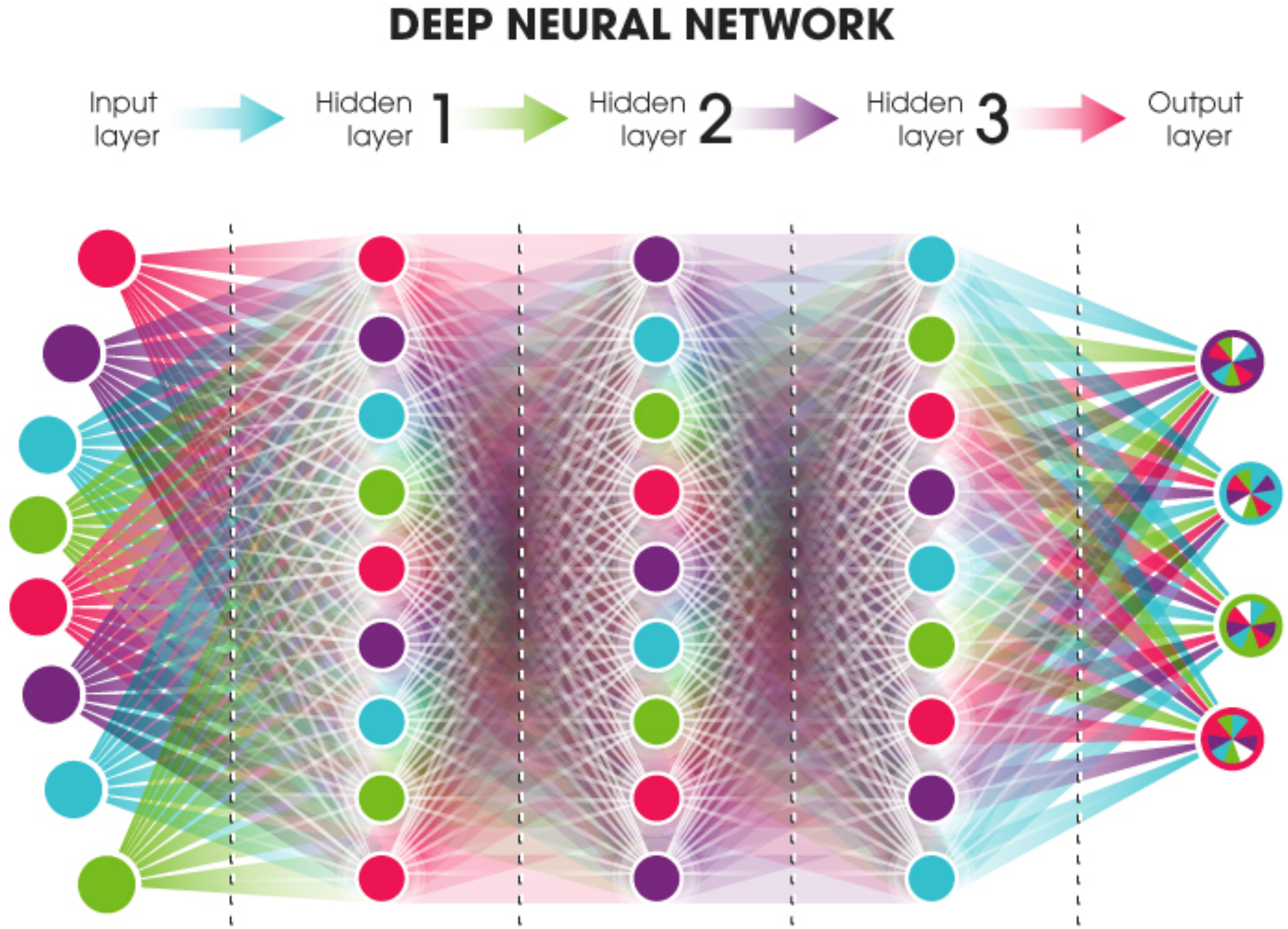
Additional bound noncoding RNA increases rate further



Making a computer out of junk



Equivalent to:



neuralnetworksanddeeplearning.com - Michael Nielsen, Yoshua Bengio, Ian Goodfellow, and Aaron Courville, 2016.

???????????????