Deep Neural Networks in Genomics

Ron Schwessinger & Edward Sanders

Hughes' Lab

MRC Molecular Haematology Unit

MRC WIMM Centre for Computational Biology

MRC Weatherall Institute of Molecular Medicine

University of Oxford

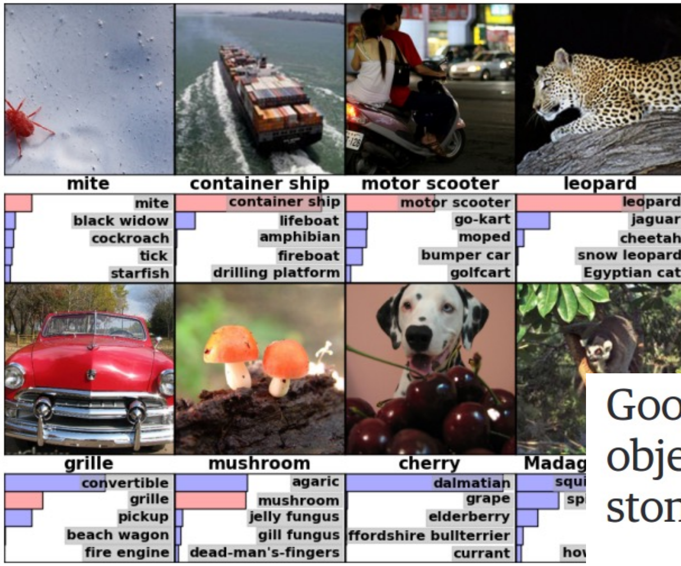
GMS teaching – 06.12.2021

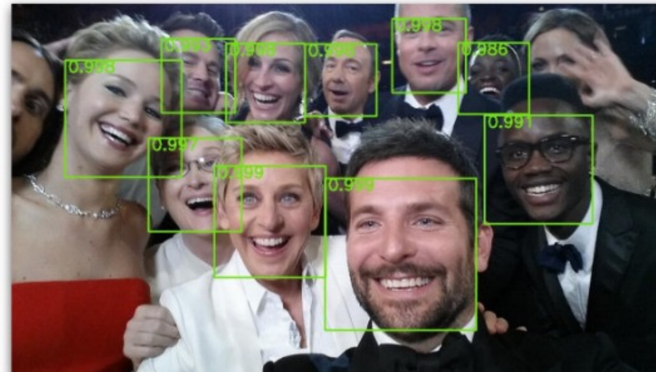
Outline

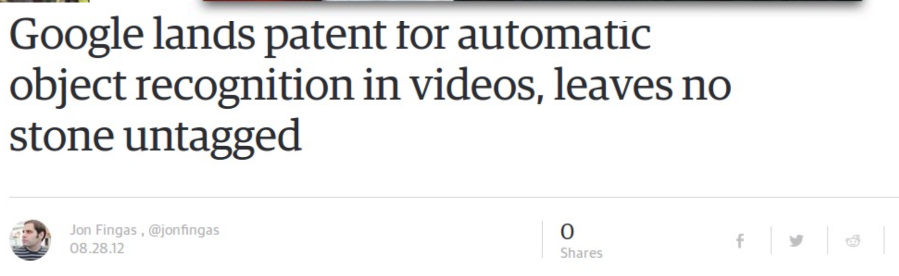
1) Introduction to Deep Neural Networks
    1) Neural Networks oversimplified
    2) Training Process
2) Introduction to Convolutional Neural Networks
    1) Convolutions for Image Analysis
    2) Convolutions for DNA
3) Convolutional Neural Networks in Genomics
    1) Chromatin Feature Networks
    2) Utility, Strategies and Interpretation
    3) More Examples, More Architectures

Introduction to Deep Neural Networks

Introduction to Deep Neural Networks



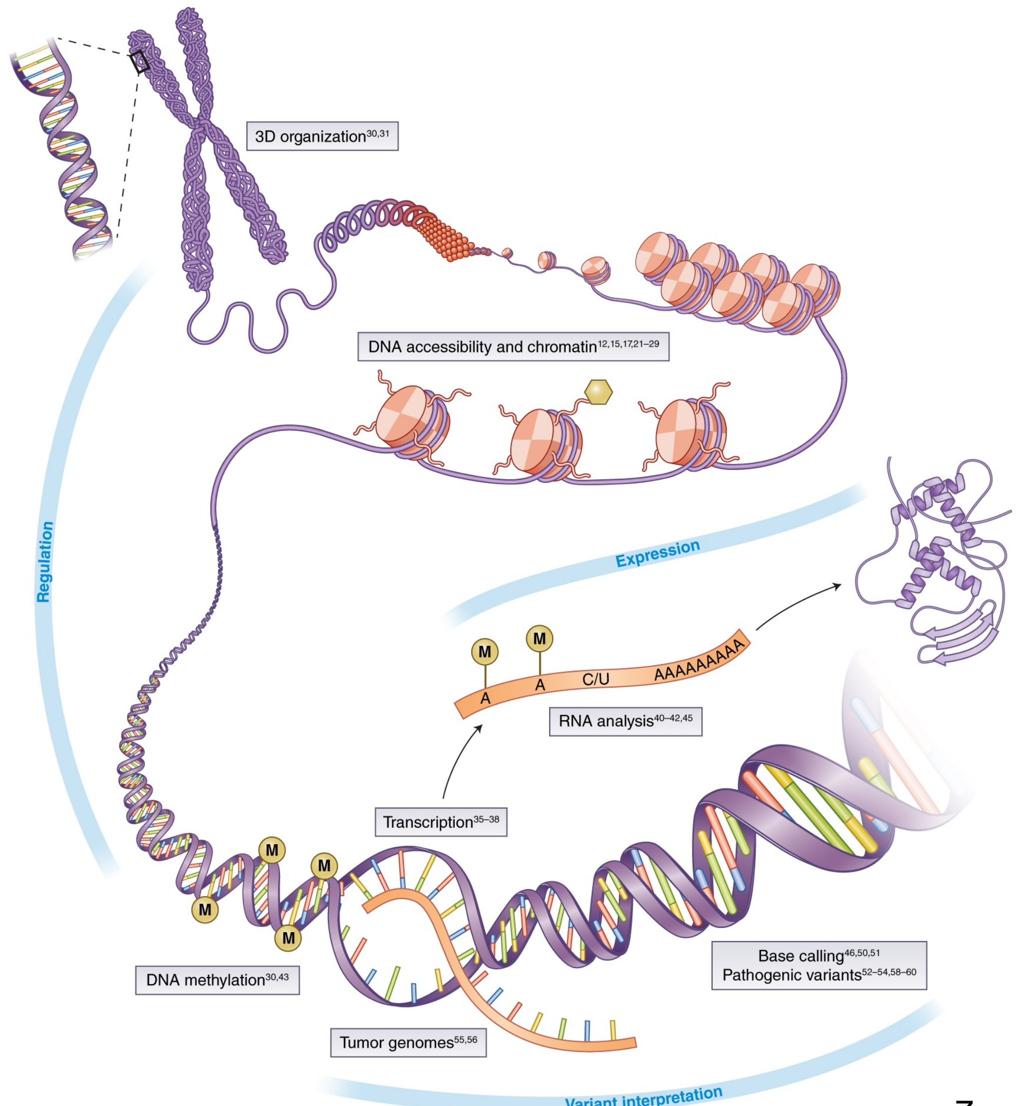



Google lands patent for automatic object recognition in videos, leaves no stone untagged

Jon Fingas , @jonfingas
08.28.12

0
Shares

Introduction to Deep Neural Networks



Zou *et al.* Nature Genetics 2019

Neural Networks oversimplified

rain probability from weather forecast x		tomorrow is a rainy day y
0.6		1
0.4	$\longrightarrow$	1
0.2		0
0.8		1

y truth: is tomorrow a rainy day

$\hat{y}$ prediction: is tomorrow a rainy day?

x local forecast rain prob.

Neural Networks oversimplified

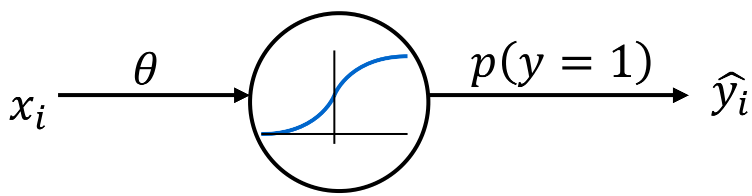
rain probability from weather forecast		tomorrow is a rainy day
x		y
0.6	$\longrightarrow$	1
0.4		1
0.2		0
0.8		1

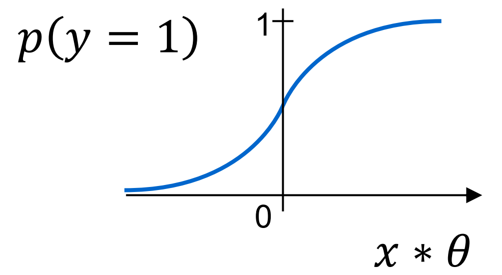
y truth: is tomorrow a rainy day

$\hat{y}$ prediction: is tomorrow a rainy day?

x local forecast rain prob.

- Single Neuron





Neural Networks oversimplified

x	w		y
0.6	0.3	UK	1
0.4	0.9	→	1
0.2	0.6		0
0.8	0.1		1

y truth is tomorrow a rainy day

$\hat{y}$ prediction: is tomorrow a rainy day?

x

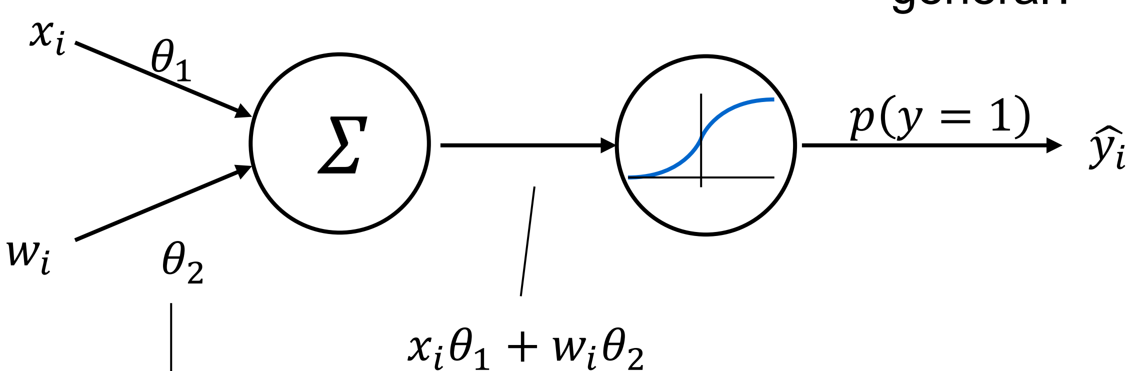
w how rainy is my current location in general?

Neural Networks oversimplified

x	w	y
0.6	0.3	1
0.4	0.9	1
0.2	0.6	0
0.8	0.1	1

UK

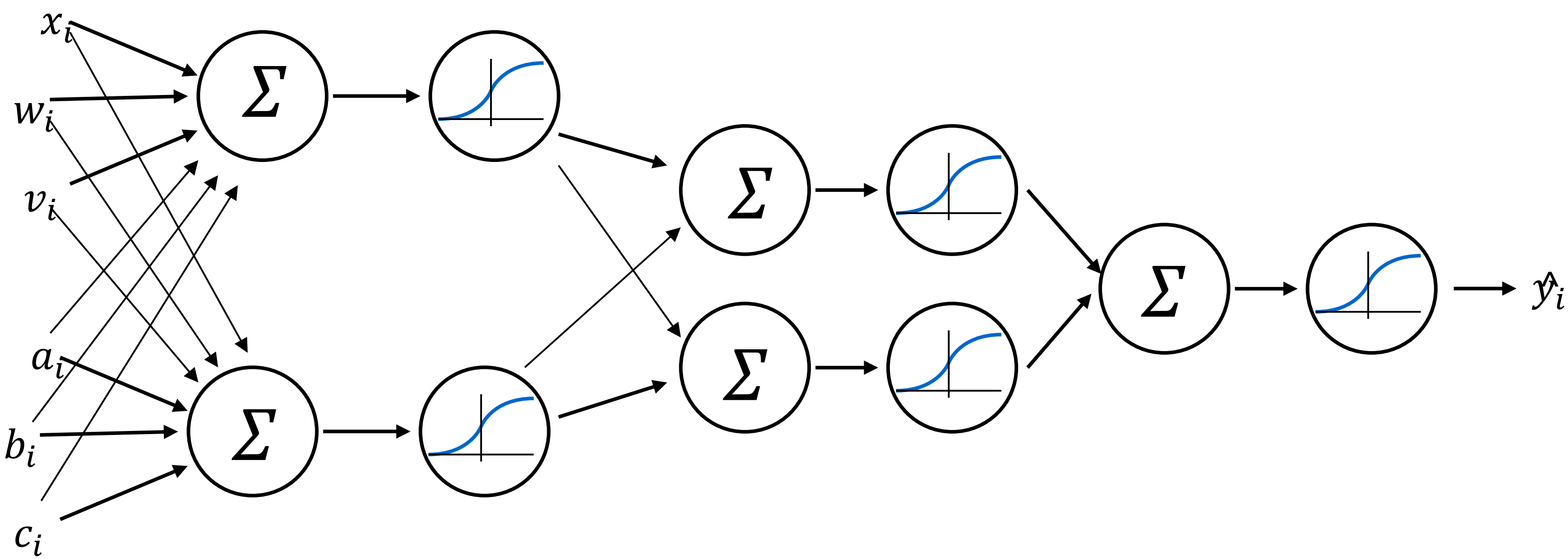
- y truth is tomorrow a rainy day
- $\hat{y}$ prediction: is tomorrow a rainy day?
- x
- w how rainy is my current location in general?



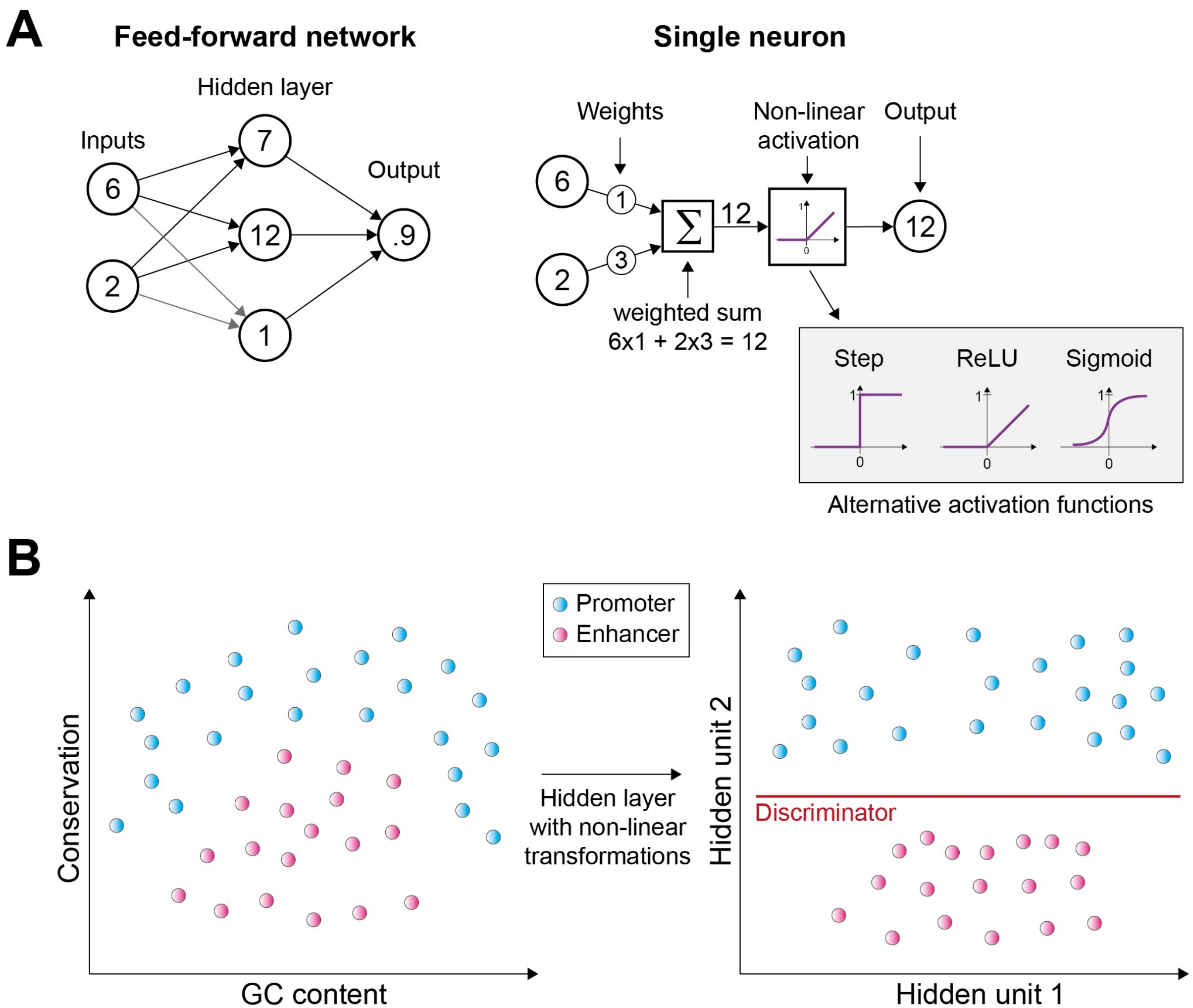
parameters

aim: tweak the parameters to get the best predictions

Neural Networks oversimplified



The Power of Non-Linear Activations



Training Process

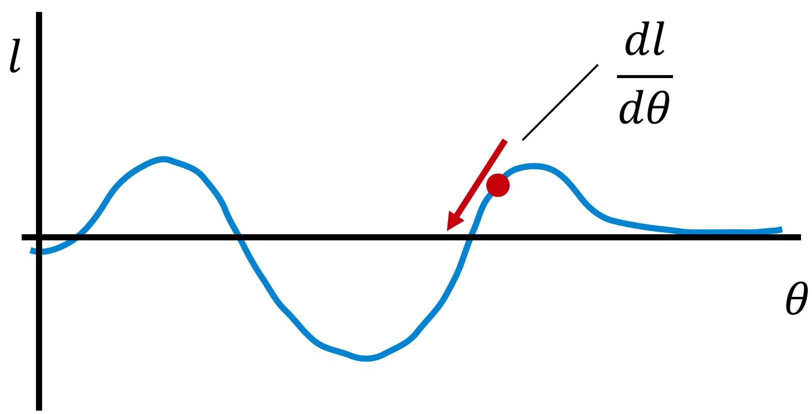
Optimise parameters θ :

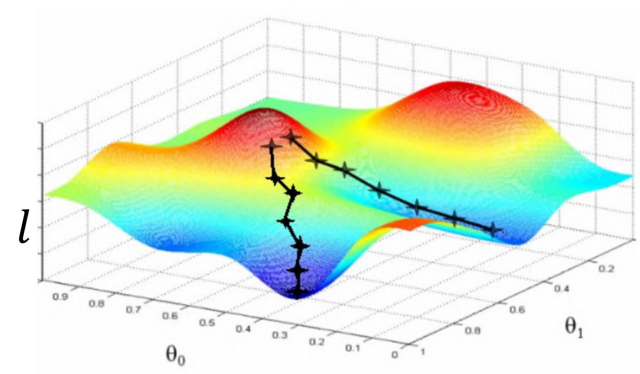
loss function l → how bad do we perfom? → minimize l

calculate **gradients!**

$$\nabla J = \begin{pmatrix} \frac{\partial l}{\partial \theta_0} \\ \frac{\partial l}{\partial \theta_1} \end{pmatrix}$$

- Optimization





Training Process

- for simple NN → gradients can be calculated by hand
- for larger architectures by computers
- but for very large/complex NNs → normal CPU computation takes ages …

- Two major developments:
  - better algorithms to calculate partial derivatives along large networks
  - using graphics cards (GPU) computation

Training Process

1) Training:

    1) Start with random parameters

    2) Take a batch of inputs

    3) Calculate predictions

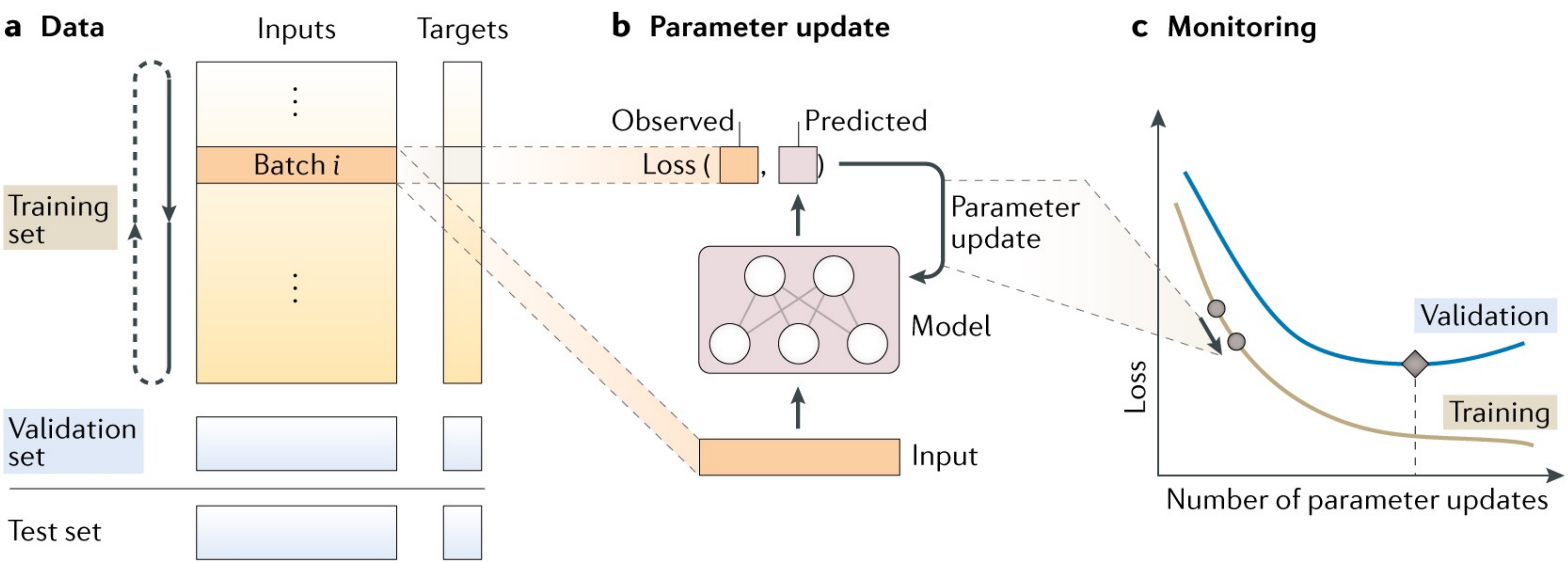
    4) Calculate your performance (loss)

    5) Calculate gradient for every parameter

    6) Adjust parameters along the gradient

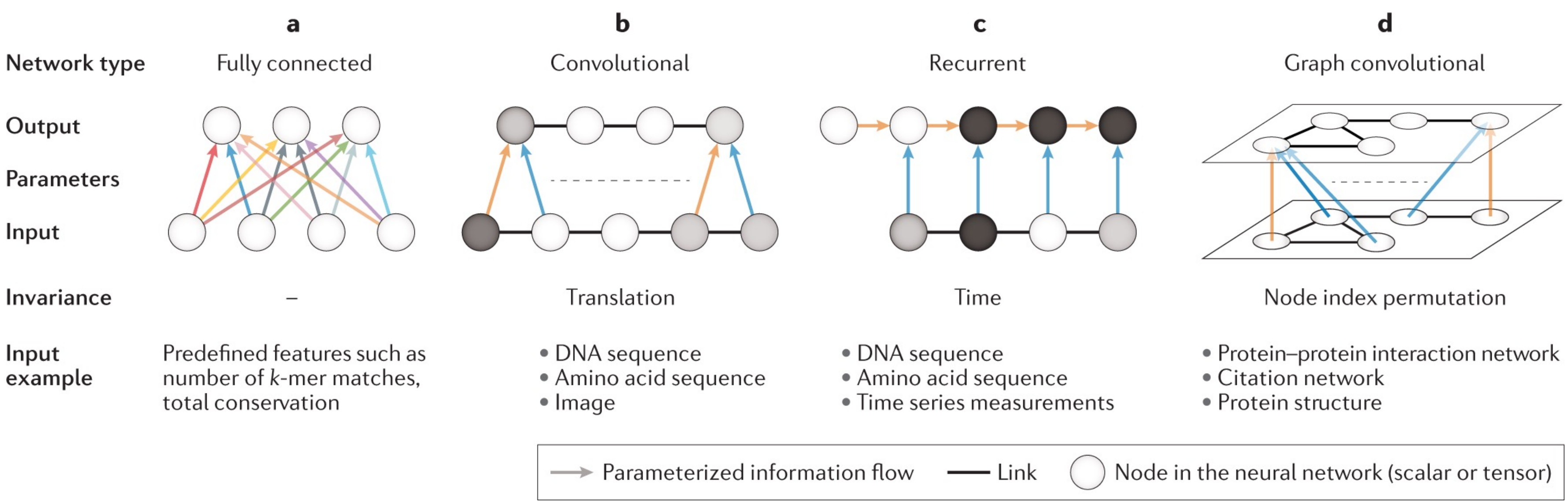
Until no improvement in performance noticeable…

Training Process



Gökcen *et al.* Nature Reviews Genetics 2019

Overview of Network Types



Gökcen *et al.* Nature Reviews Genetics 2019

Introduction to Convolutional Neural Networks

Convolutions for Image Analysis

Image

Grey values per pixels

104	111	101	101
90	88	91	90
39	41	35	44
51	49	48	52

Convolutions for Image Analysis

Image

Grey values per pixels

104	111	101	101
90	88	91	90
39	41	35	44
51	49	48	52

$\rightarrow$

192	202	191
131	123	135
88	89	87

Filter

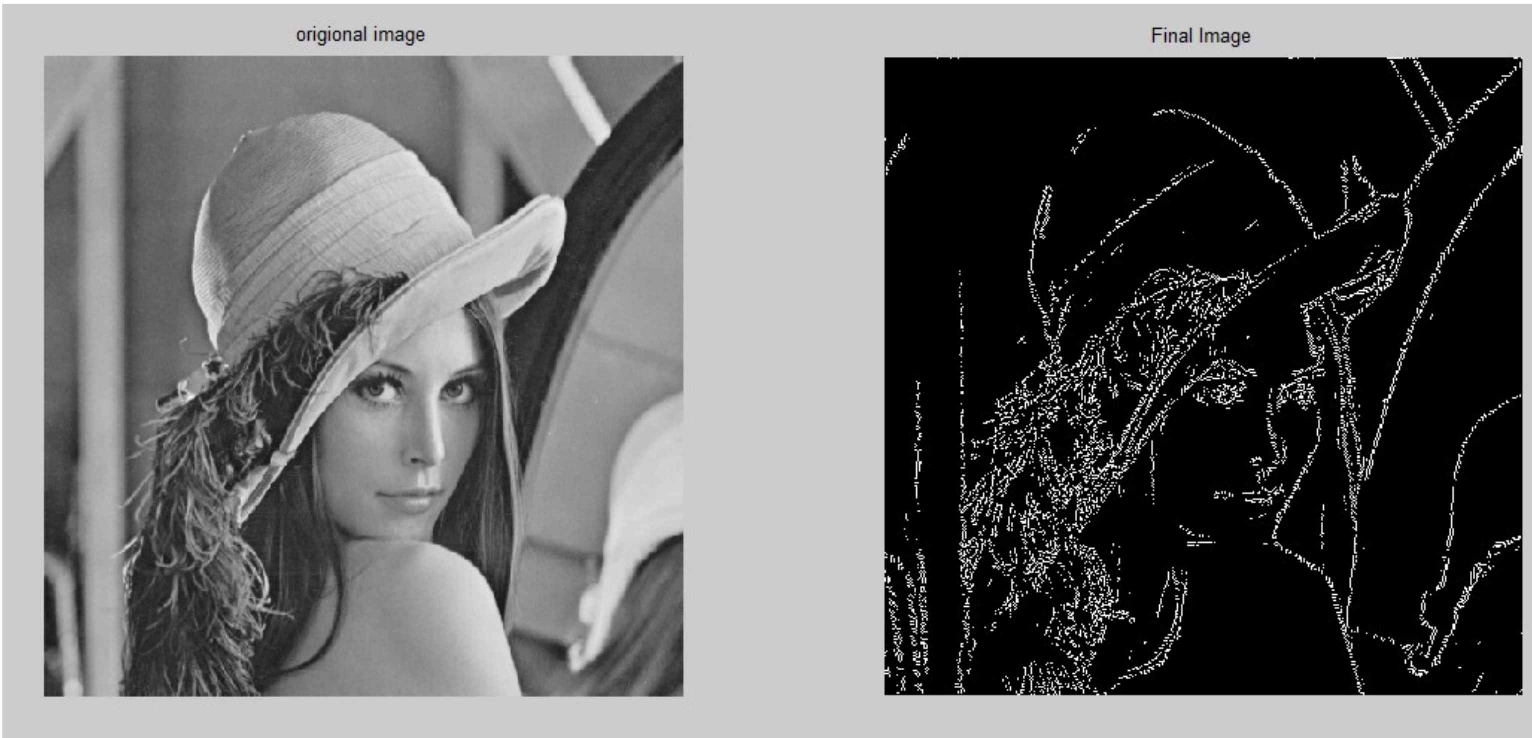
$$\Sigma\left(\begin{matrix}1 & 0\\ 0 & 1\end{matrix} * \begin{matrix}a & b\\ c & d\end{matrix}\right) = \mathbf{x}$$

$$1*a + 0*b + 0*c + 1*d = x$$

$$\text{Filter} = \begin{matrix}\theta_1 & \theta_2\\ \theta_3 & \theta_4\end{matrix} \rightarrow \text{parameters to optimize}$$

Convolutions for Image Analysis

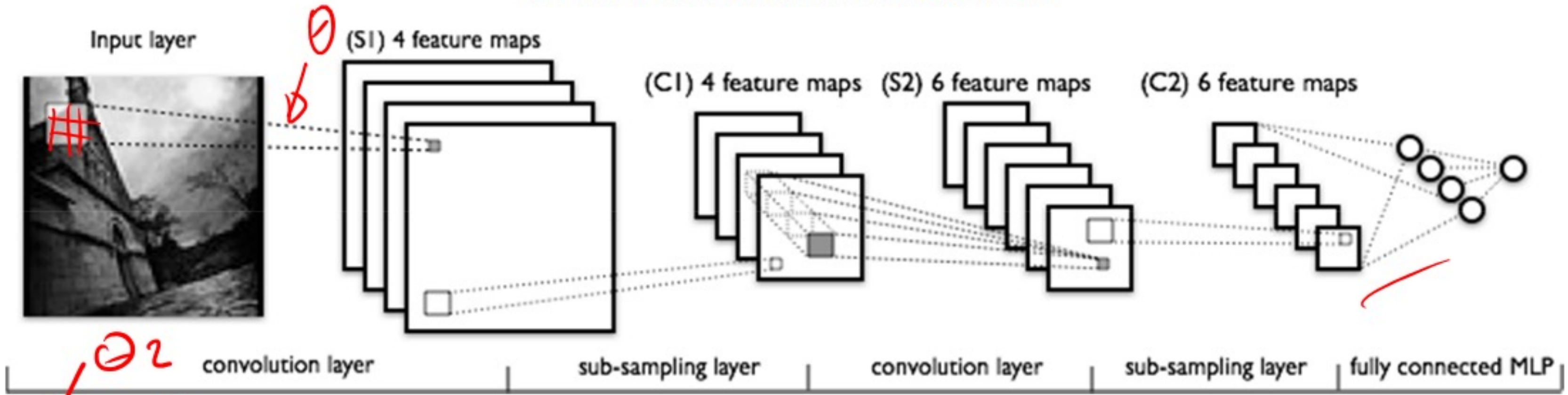
Edge detection – Sobel operator

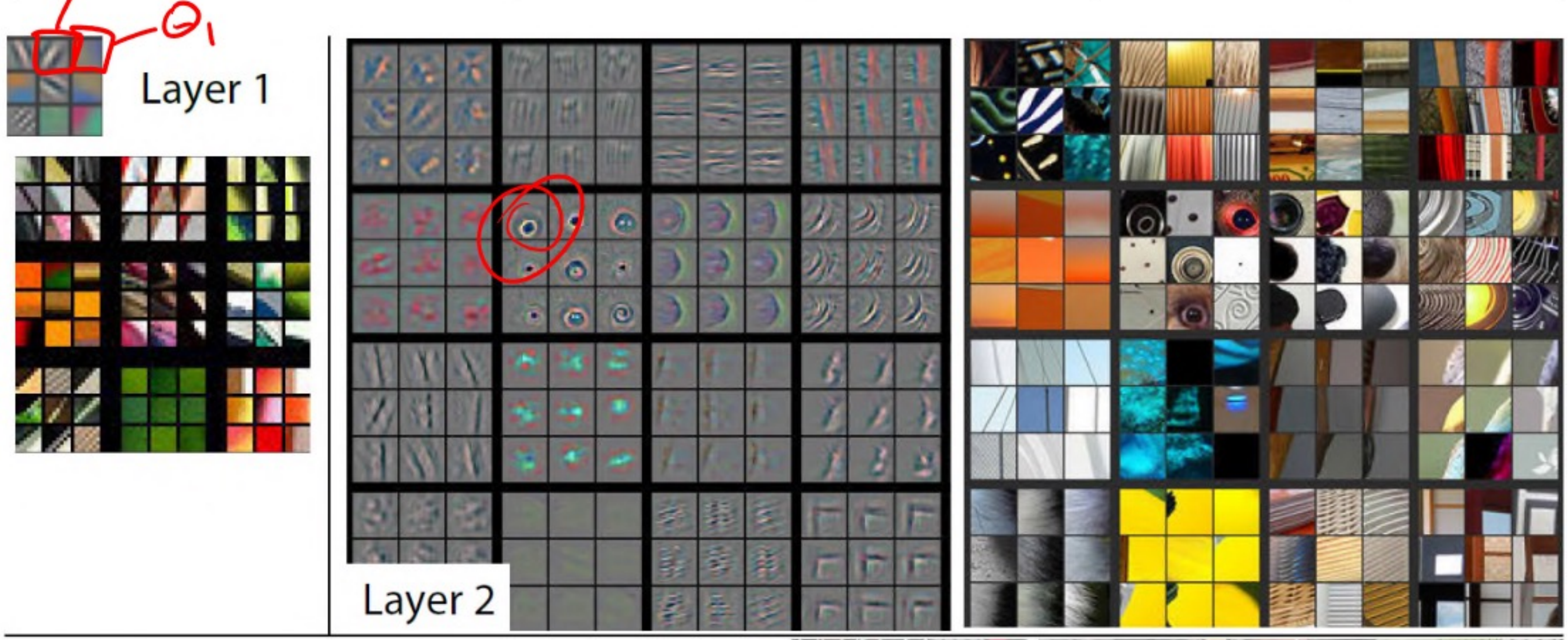


-1	0	1
-2	0	2
-1	0	1

1	2	1
0	0	0
-1	-2	-1

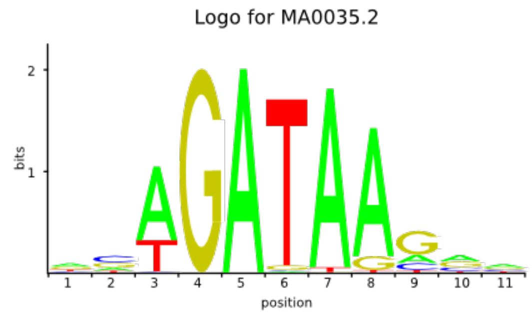
Convolutional networks





[Matthew Zeiler & Rob Fergus]

Convolutions for DNA

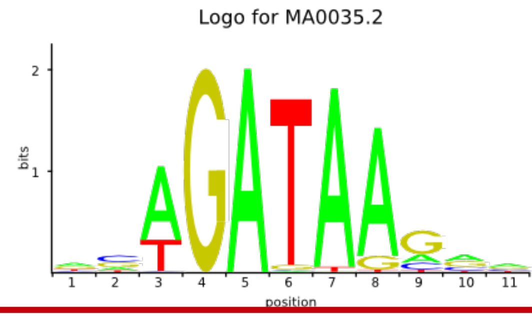


A [1423 708 2782 0 4000 27 3887 3550 799 1432 1487]
C [560 1633 31 0 0 29 0 4 681 897 829]
G [1242 1235 10 4000 0 109 6 383 2296 1360 1099]
T [775 424 1177 0 0 3835 107 63 224 311 585]

hot coded sequence

A C A G A T A A G T A G A G G C T A T T C C …
A [1 0 1 0 1 0 1 1 0 0 1 0 1 0 0 0 0 1 0 0 0 0 …]
C [0 1 0 0 0 0 0 0 0 0 0 0 0 0 0 1 0 0 0 0 1 1 …]
G [0 0 0 1 0 0 0 0 1 0 0 1 0 1 1 0 0 0 0 0 0 0 …]
T [0 0 0 0 0 1 0 0 0 1 0 0 0 0 0 0 1 0 1 1 0 0 …]

Convolutions for DNA



A [1423 708 2782 0 4000 27 3887 3550 799 1432 1487]
C [560 1633 31 0 0 29 0 4 681 897 829]
G [1242 1235 10 4000 0 109 6 383 2296 1360 1099]
T [775 424 1177 0 0 3835 107 63 224 311 585]

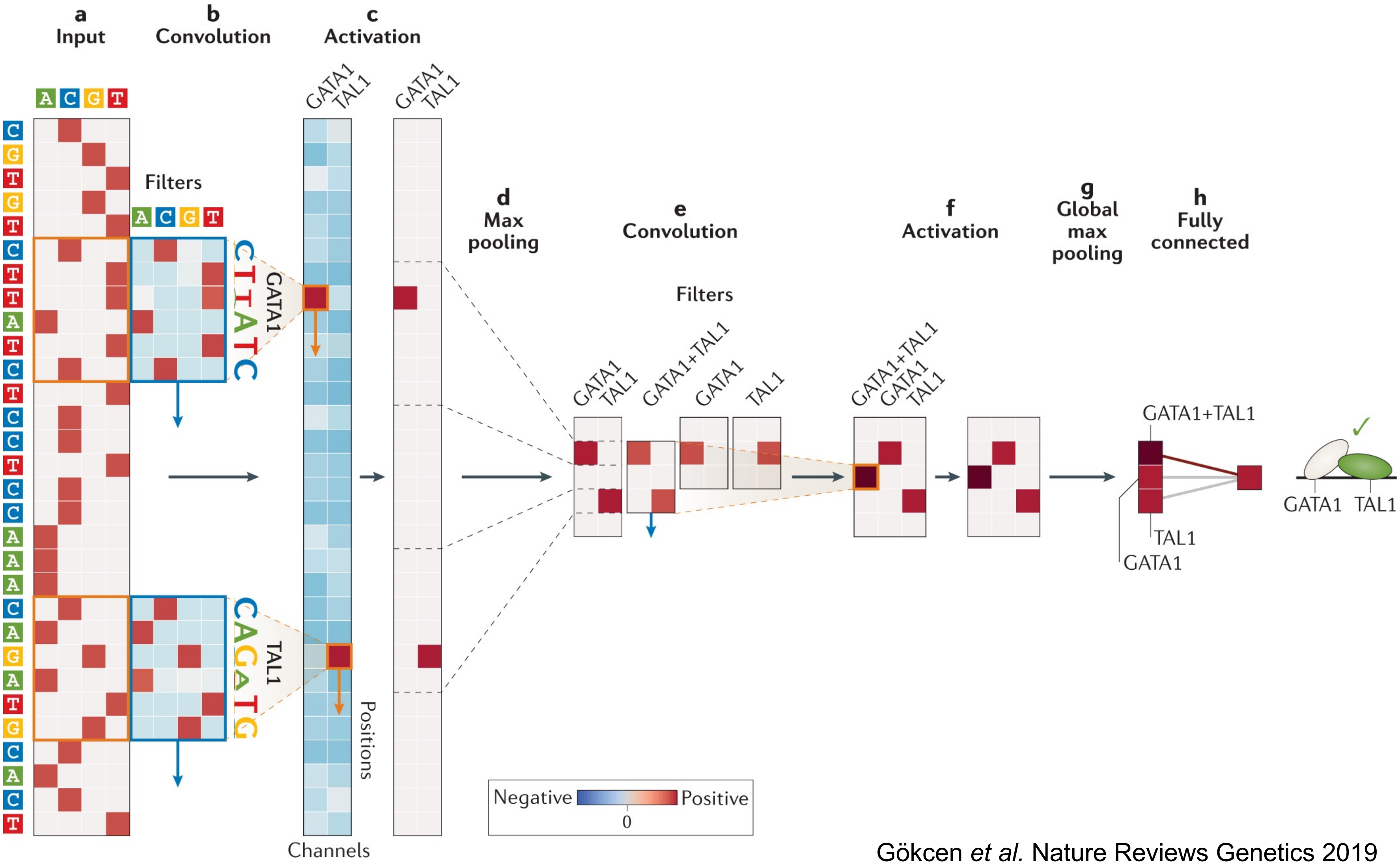
hot coded sequence

A C A G A T A A G T A G A G G C T A T T C C ...

A [1 0 1 0 1 0 1 1 0 0 1 0 1 0 0 0 0 1 0 0 0 0 ...]
C [0 1 0 0 0 0 0 0 0 0 0 0 0 0 0 1 0 0 0 0 1 1 ...]
G [0 0 0 1 0 0 0 0 1 0 0 1 0 1 1 0 0 0 0 0 0 0 ...]
T [0 0 0 0 0 1 0 0 0 1 0 0 0 0 0 0 1 0 1 1 0 0 ...]

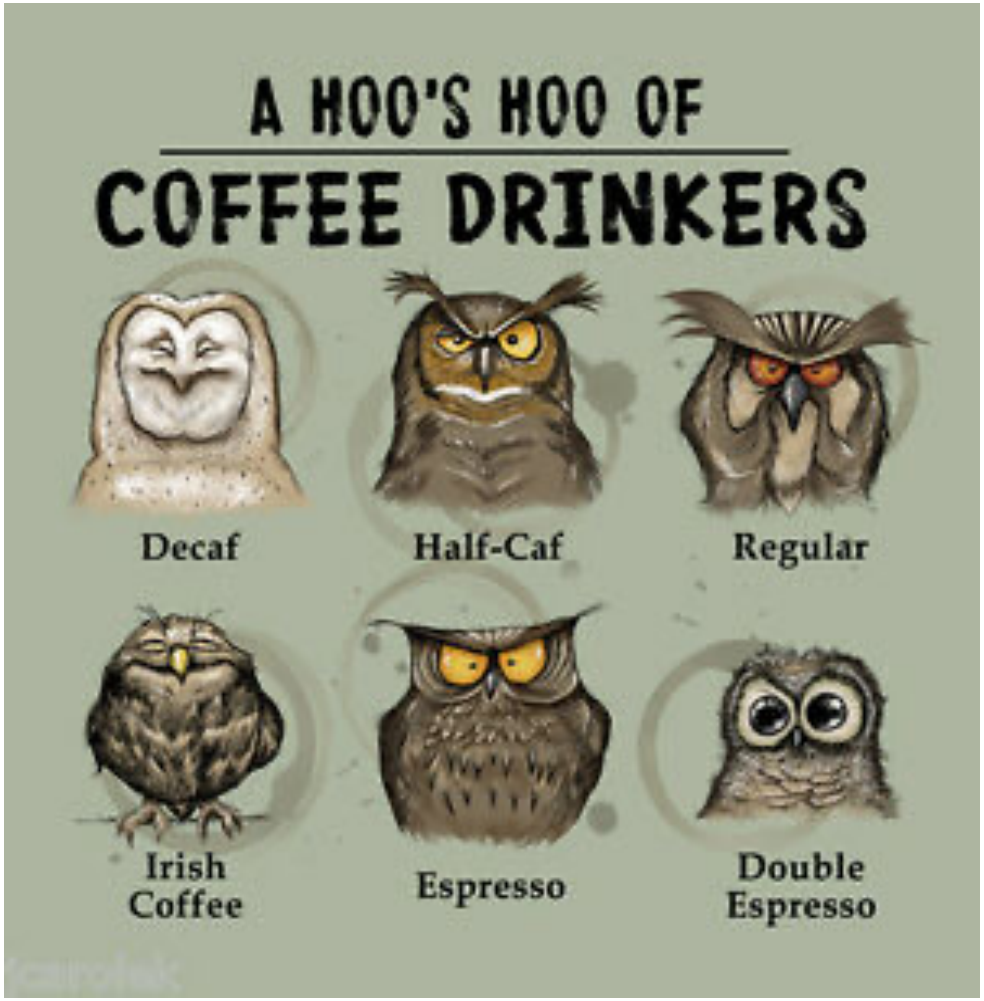
1005 506 501 501 40 55 501 ...

Convolutions for DNA

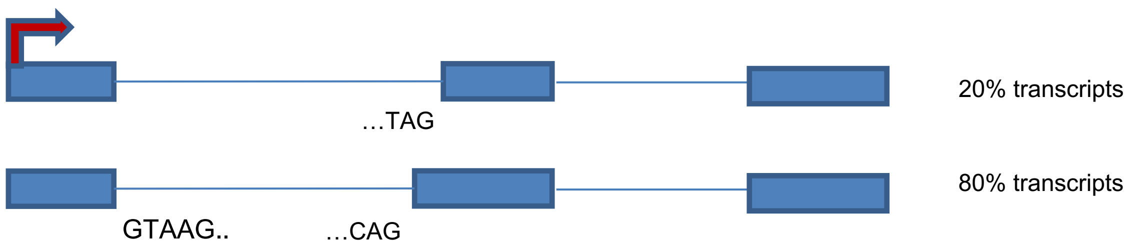


Gökcen *et al.* Nature Reviews Genetics 2019

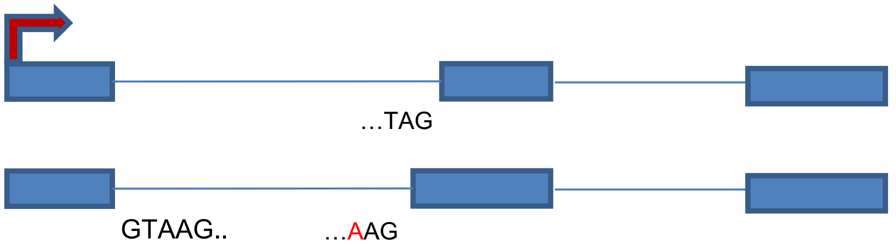
Coffee Break!



Transcript exon prediction



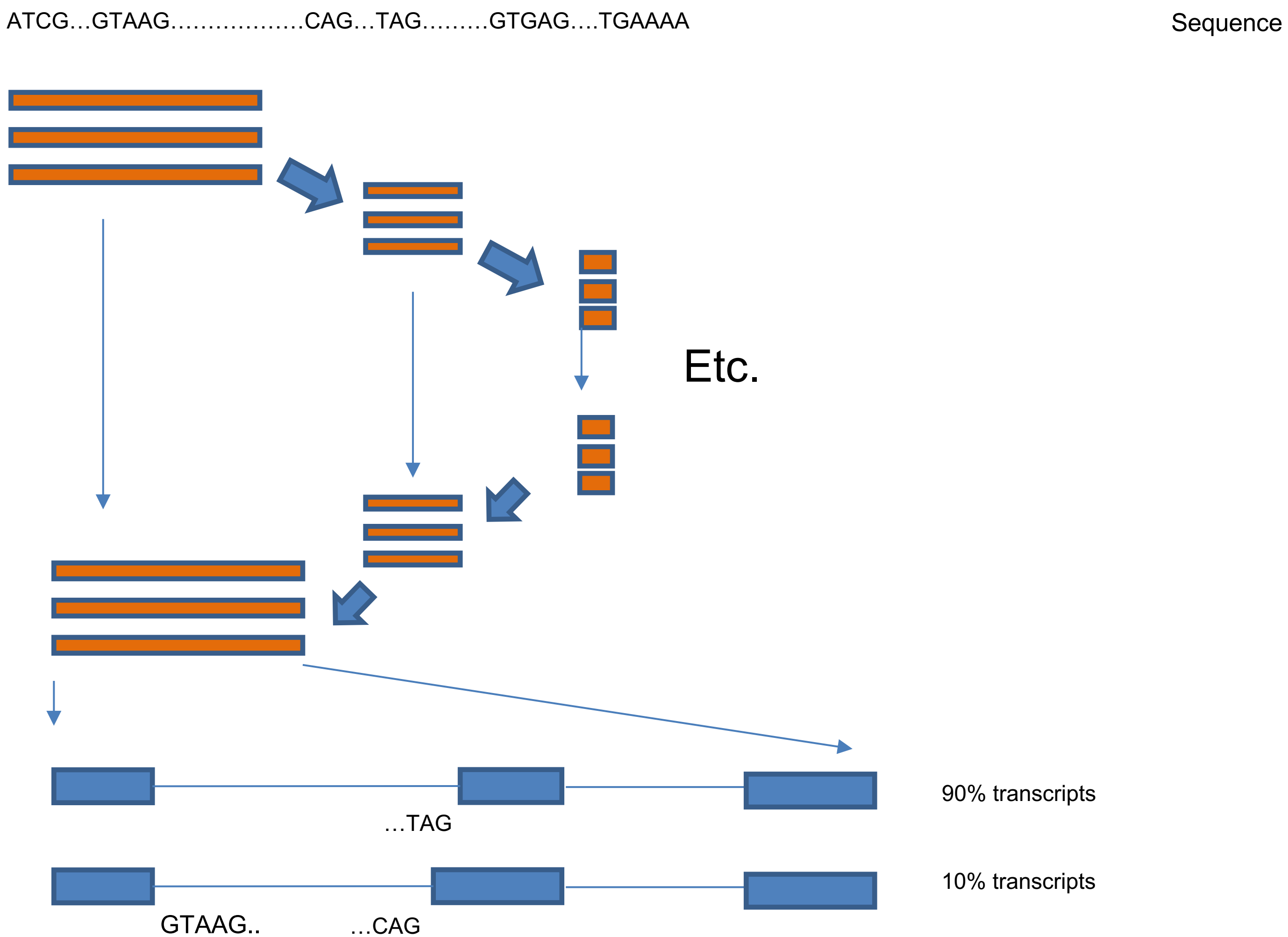
Transcript exon prediction



90% transcripts

10% transcripts

Transcript exon prediction



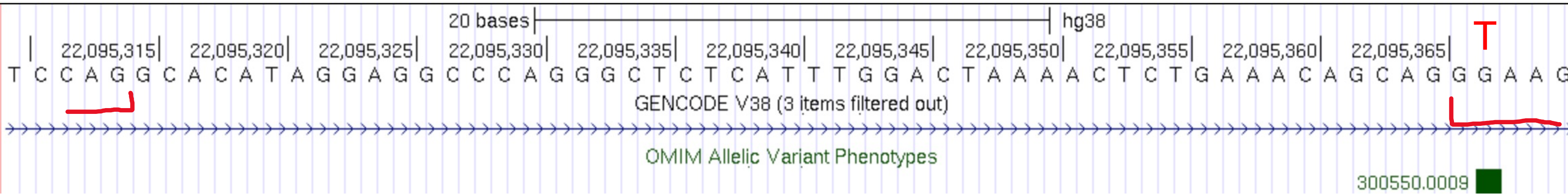
Transcript exon prediction -- application to mutagenesis

Trains function f where f(sequence) outputs probability distribution of ”x being a donor/acceptor” at genome positions

Can run:

1. **f**(sequence)
2. **f**(sequence + mutation)

Transcript exon prediction -- Application to disease diagnostics

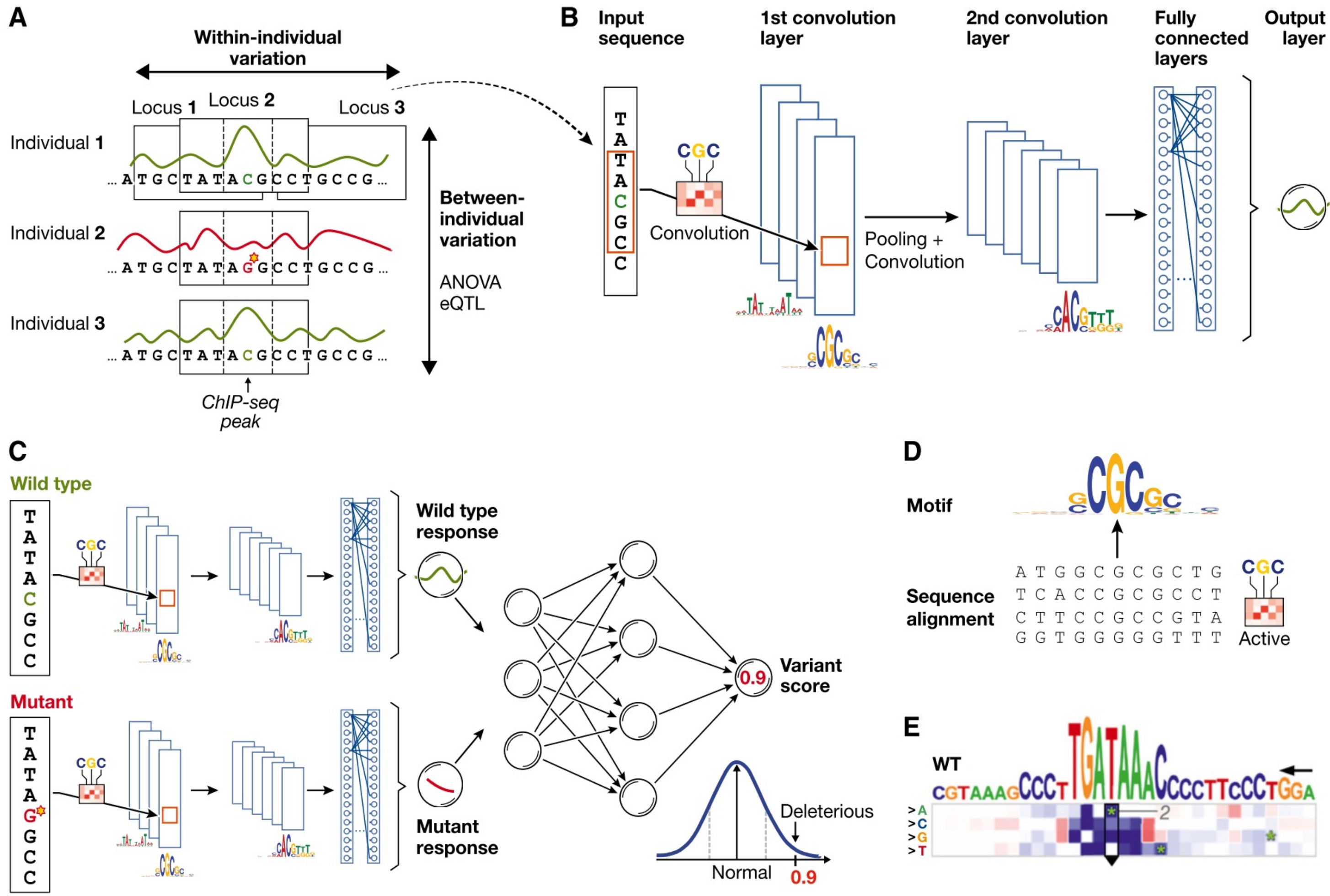


Apply model with G -> T mutation

Model predicts inclusion of novel exon replicating known result in OMIN.

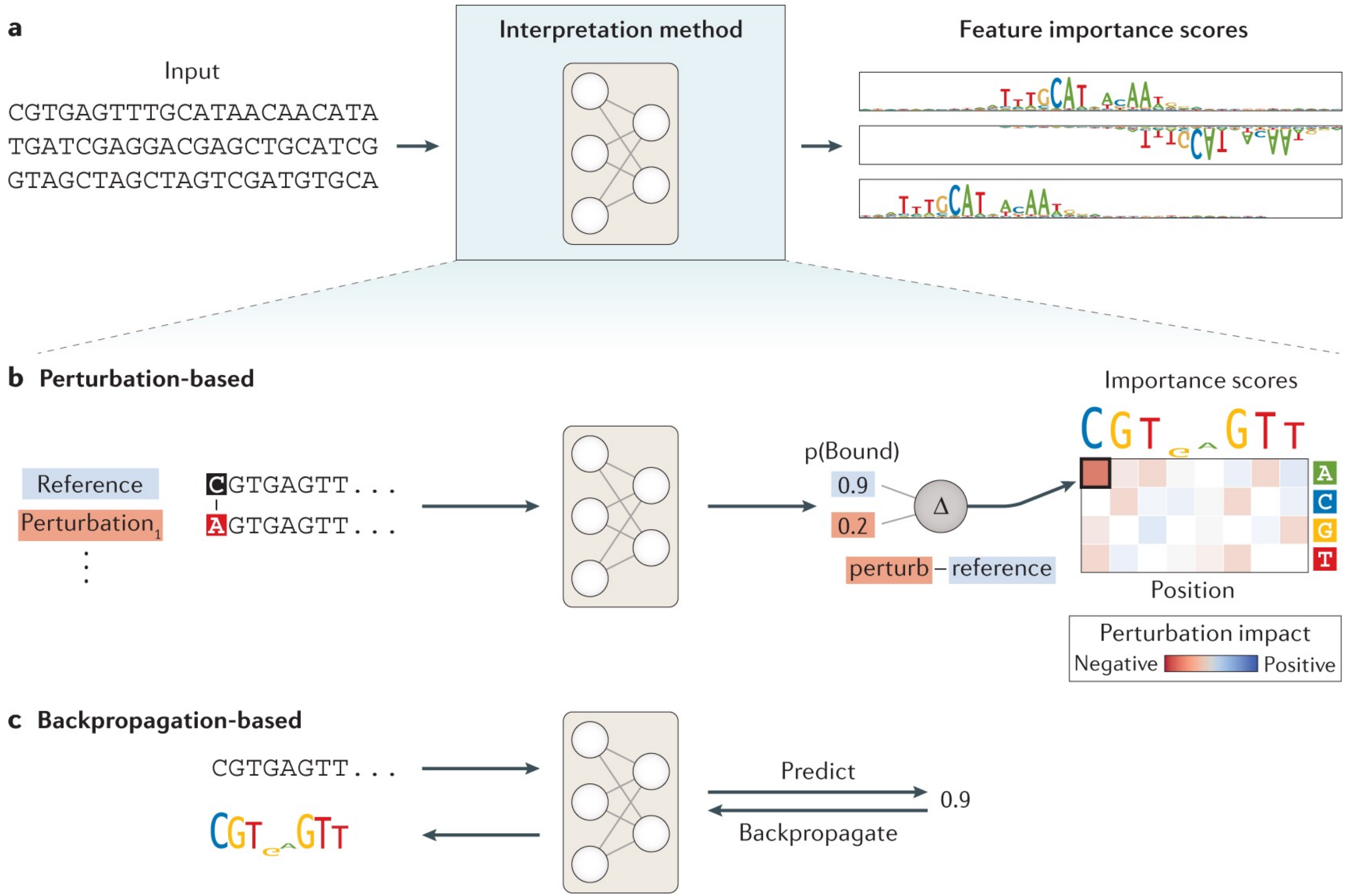
In fact -- such models act as precise predictors for known splice-altering disease causing variants.

Utility



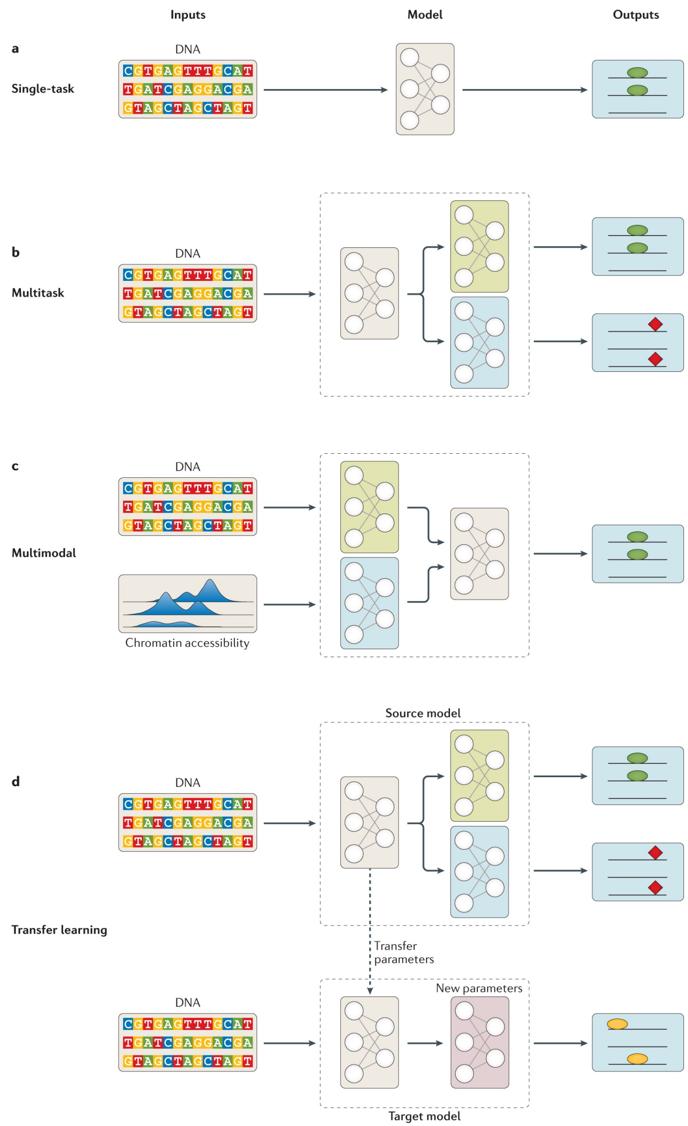
Angermueller *et al.* 2016

Interpretation



Gökcen *et al.* Nature Reviews Genetics 2019

Learning Strategy

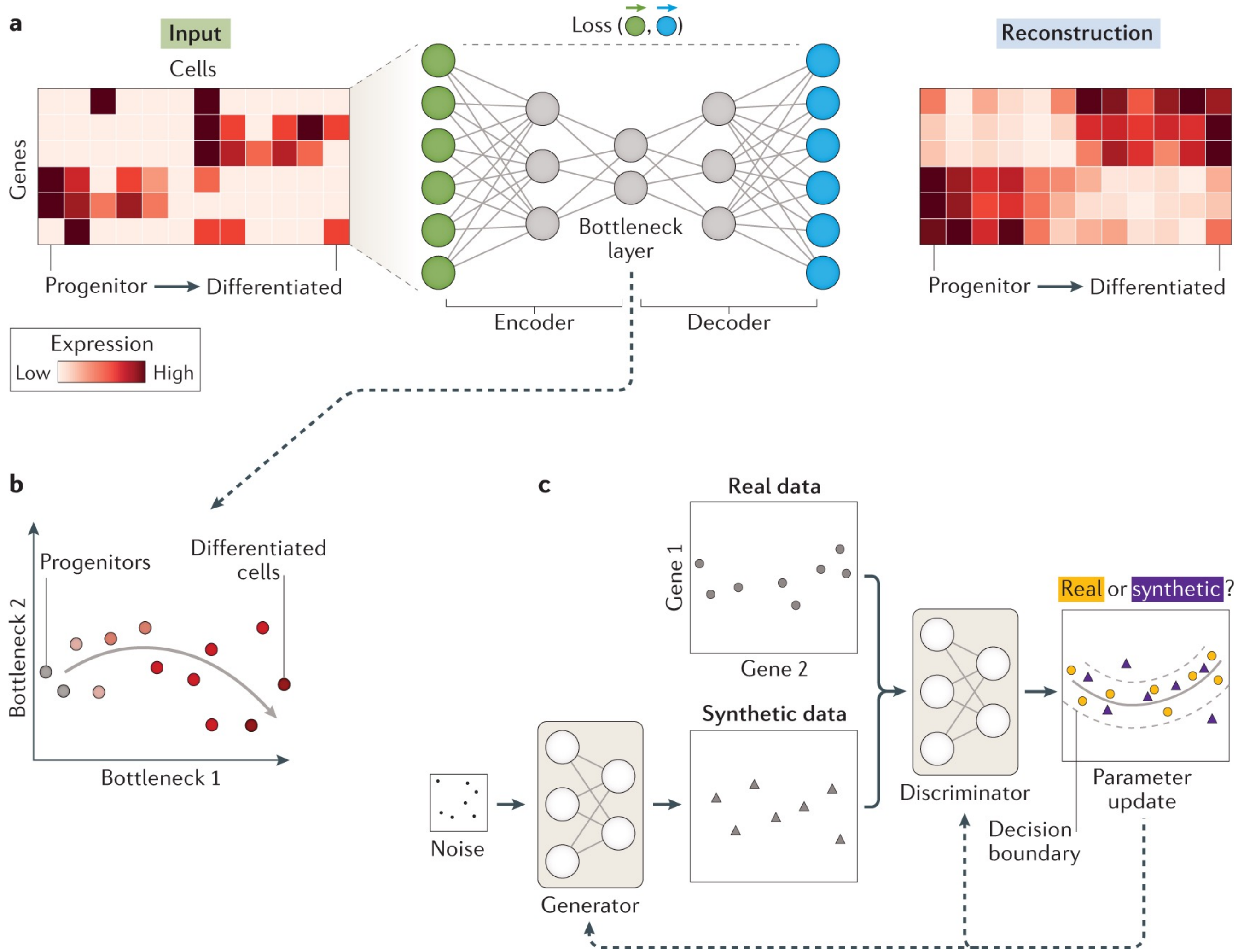


Gökcen *et al.* Nature Reviews Genetics 2019

- More Examples, More Architectures

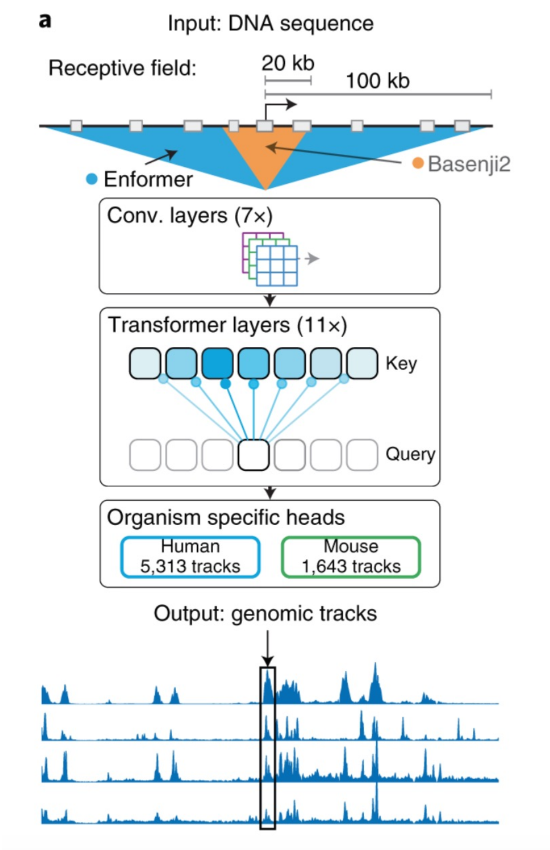
Gökcen *et al.* Nature Reviews Genetics 2019

AutoEncoders and GANs



Gökcen *et al.* Nature Reviews Genetics 2019

Attention networks & Enformer model for gene expression



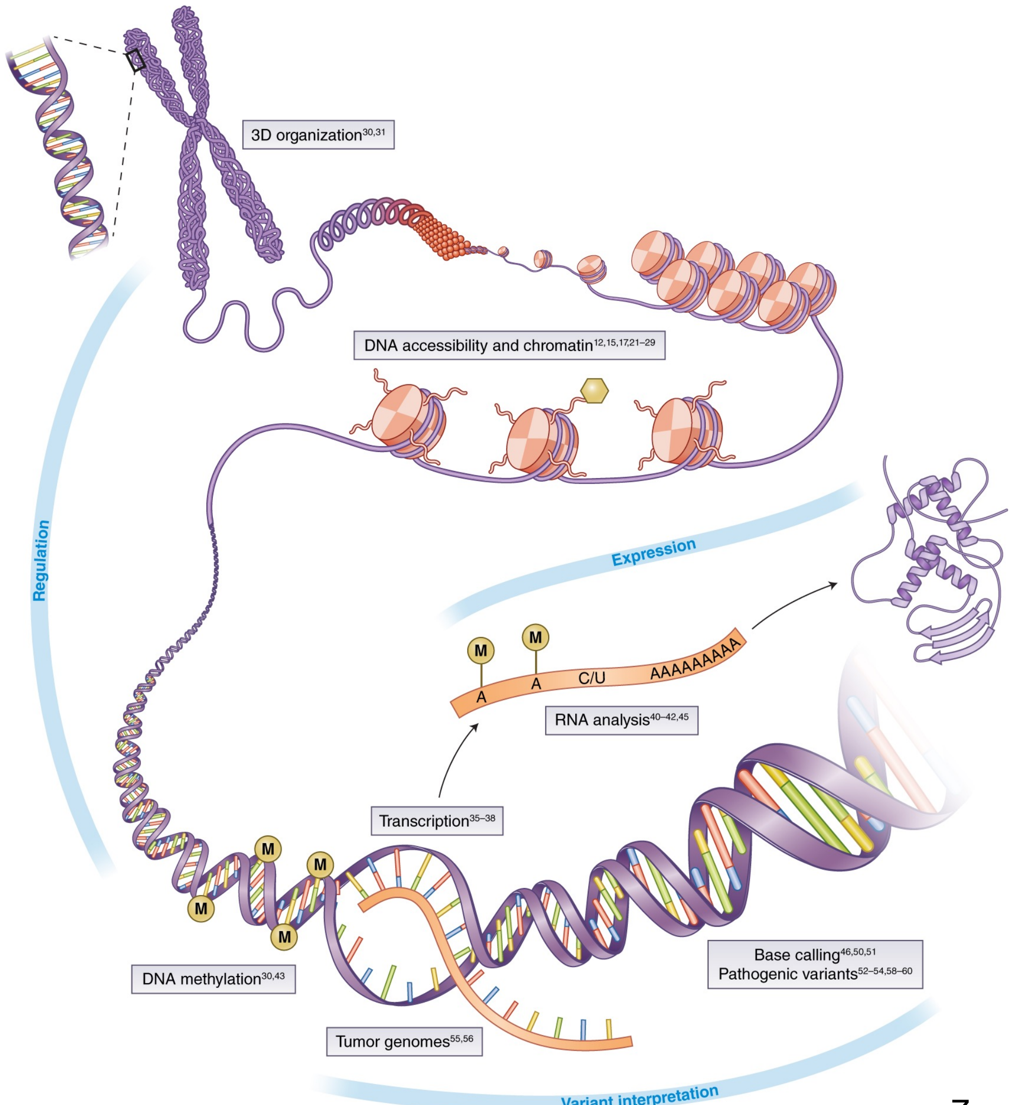
Attention networks explicitly model (potentially long range) interactions.

Similar to “YouTube” -- it learns “queries”, “keys” and “values”. All come from the genetic sequence.

Disadvantage (for us) coming from model size

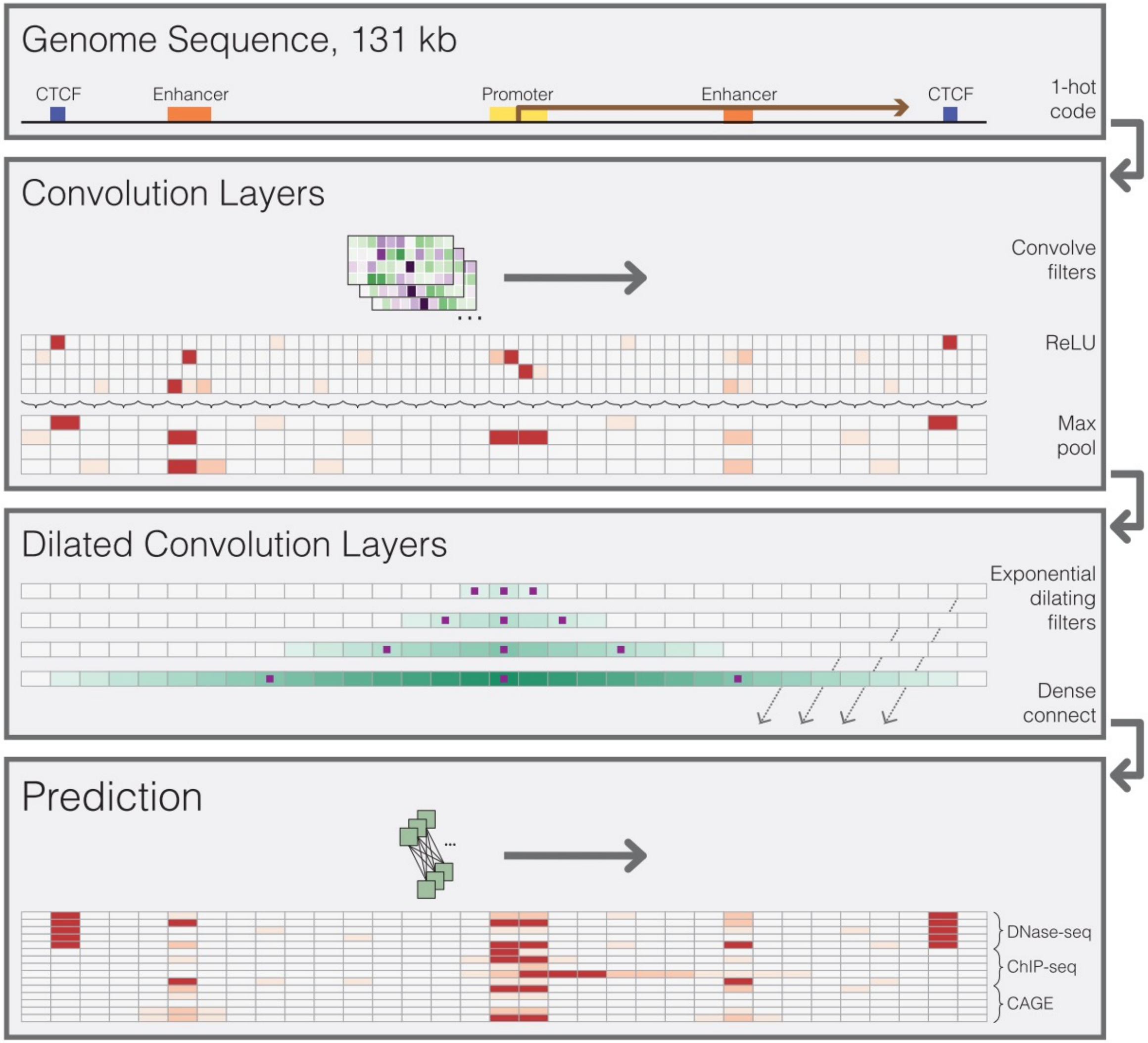
Kelley *et al.* Nature Methods 2021

Deep Learning in Genomics



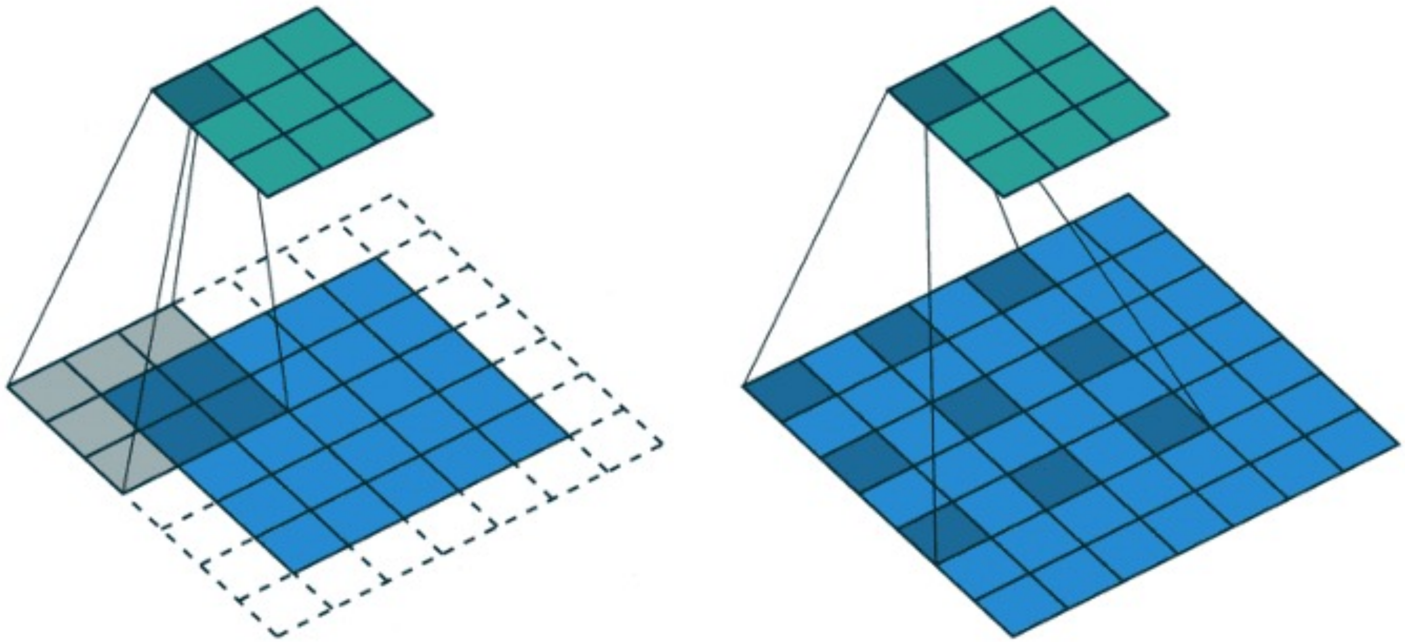
Zou *et al.* Nature Genetics 2019

Examples – Dilated Networks

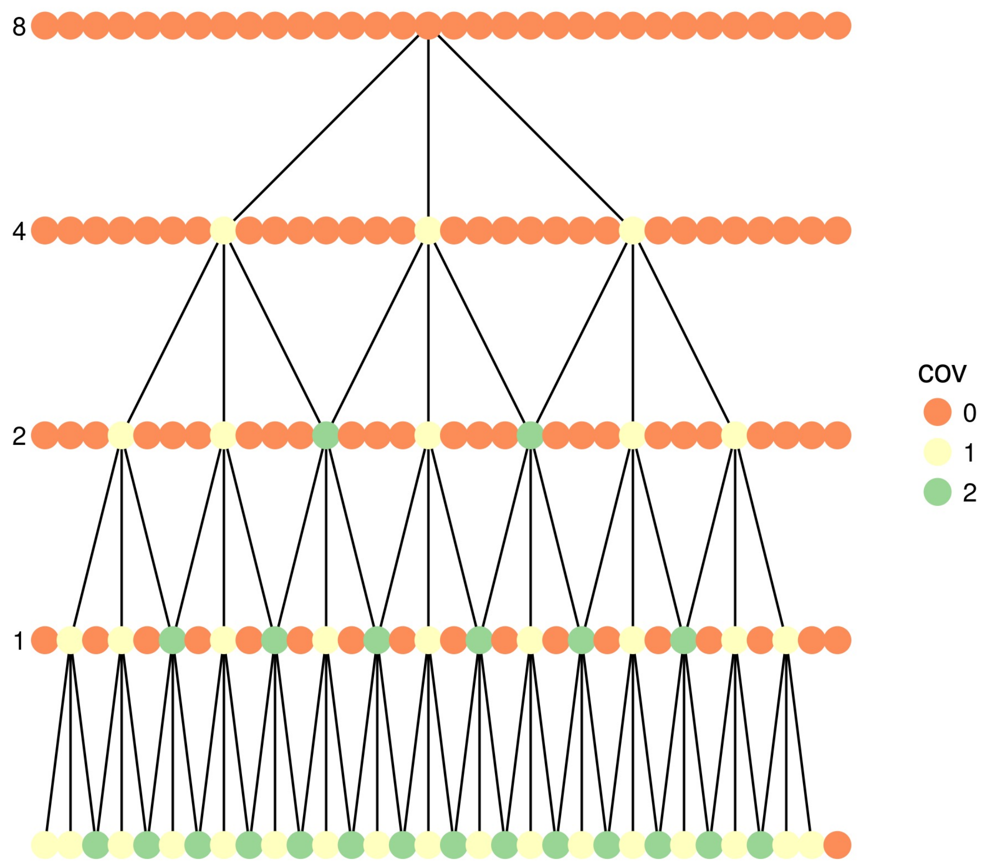


Kelley *et al.* Genome Research 2018

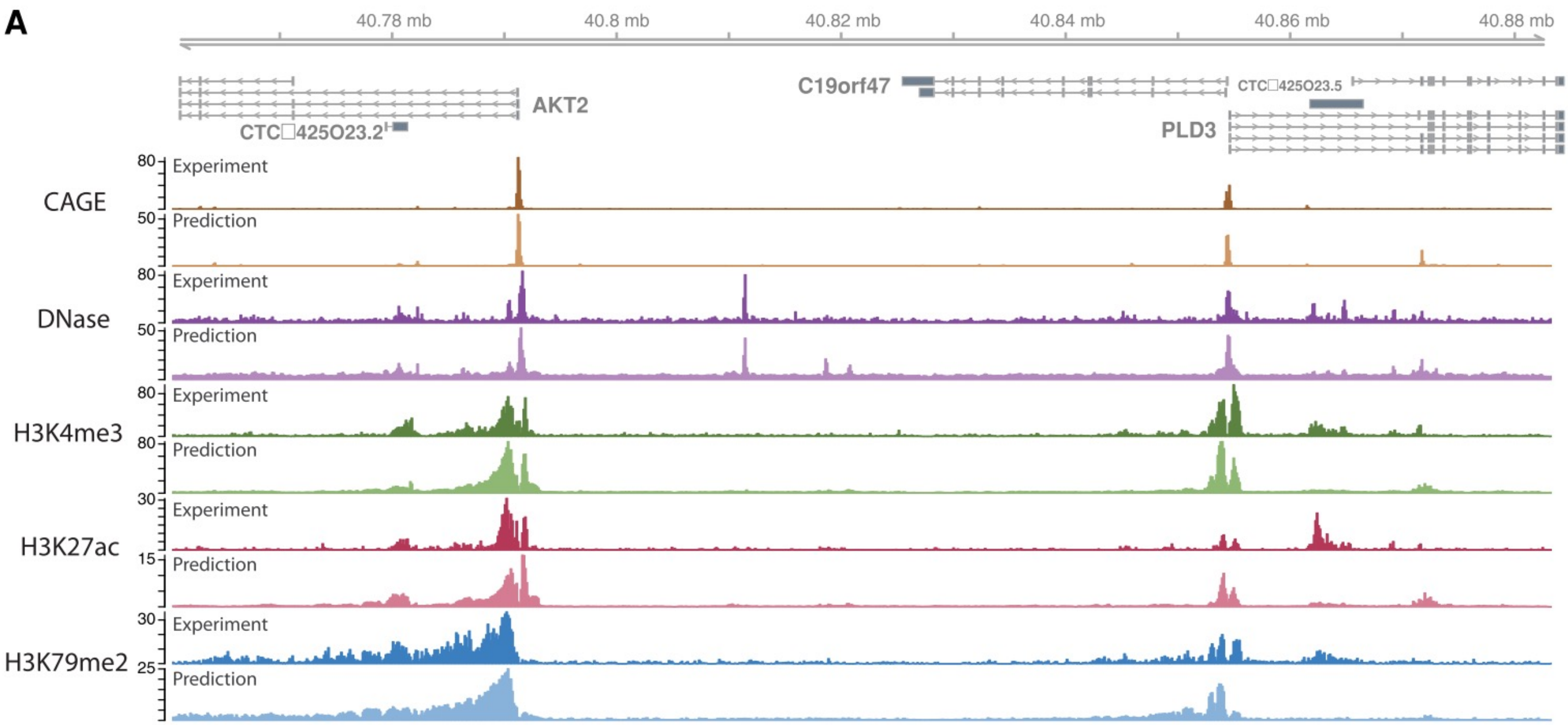
Examples – Dilated Networks



Examples – Dilated Networks

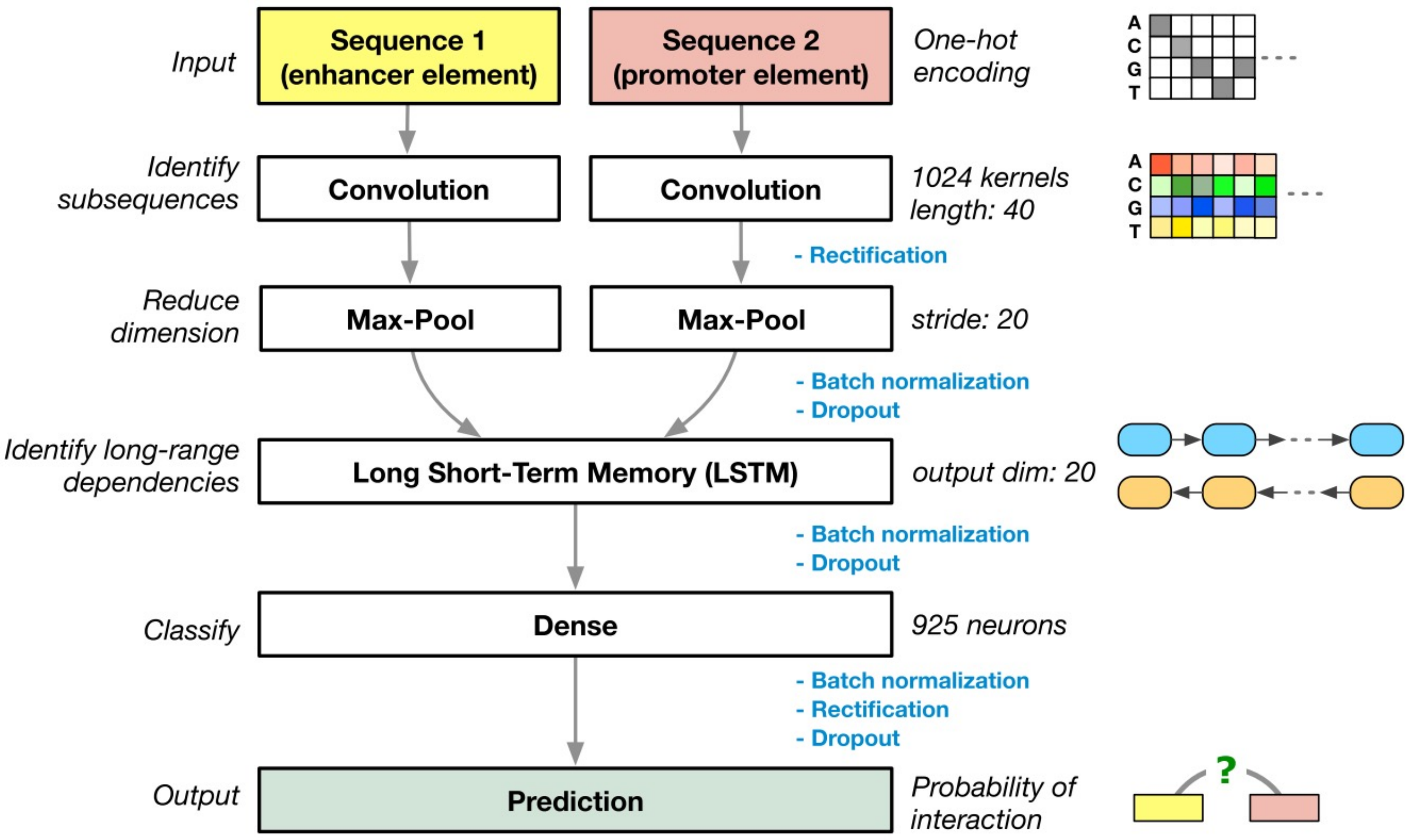


Examples – Dilated Networks



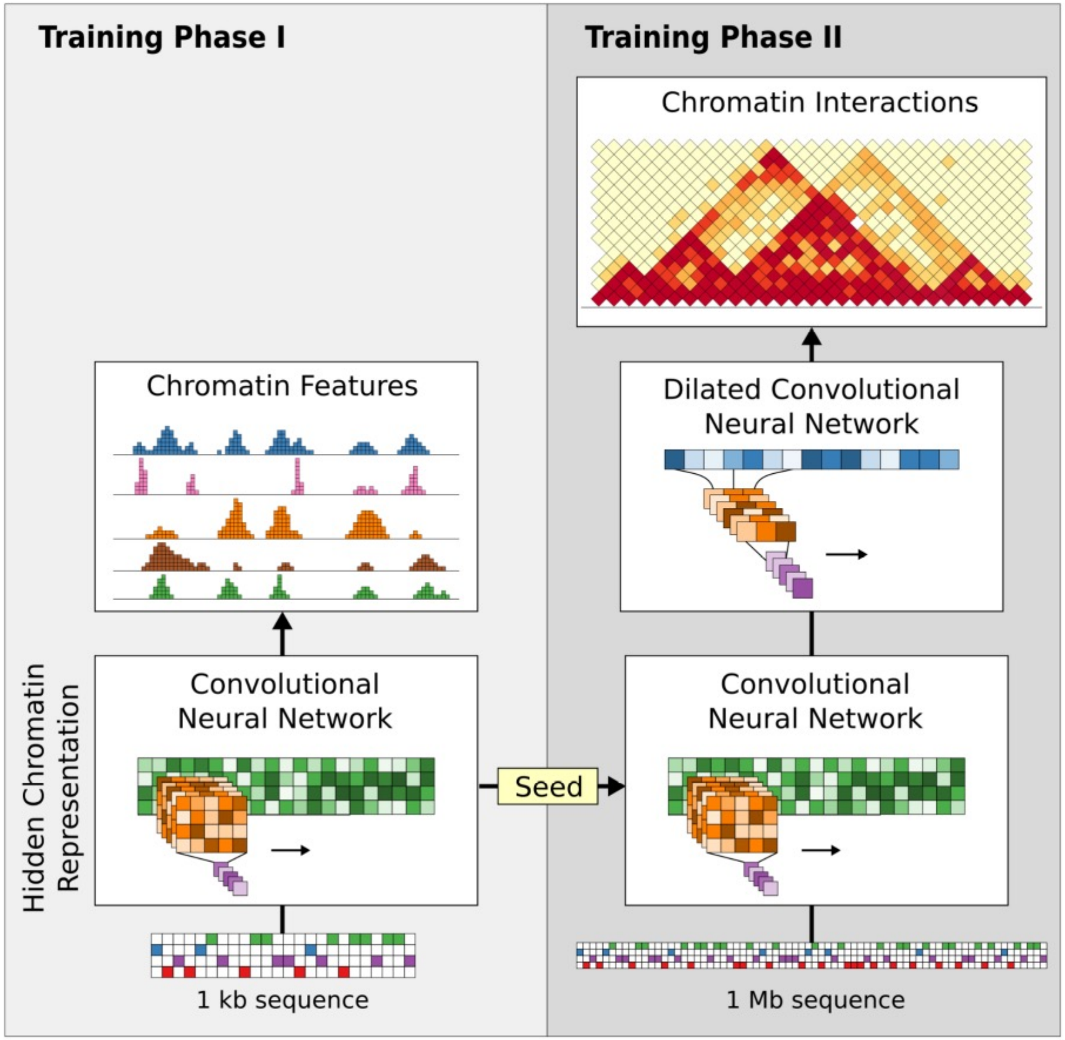
Kelley *et al.* Genome Research 2018

Examples – Enhancer – Promoter - Contacts



Singh *et al.* Quantitative Biology 2019

Examples – 3D Genome

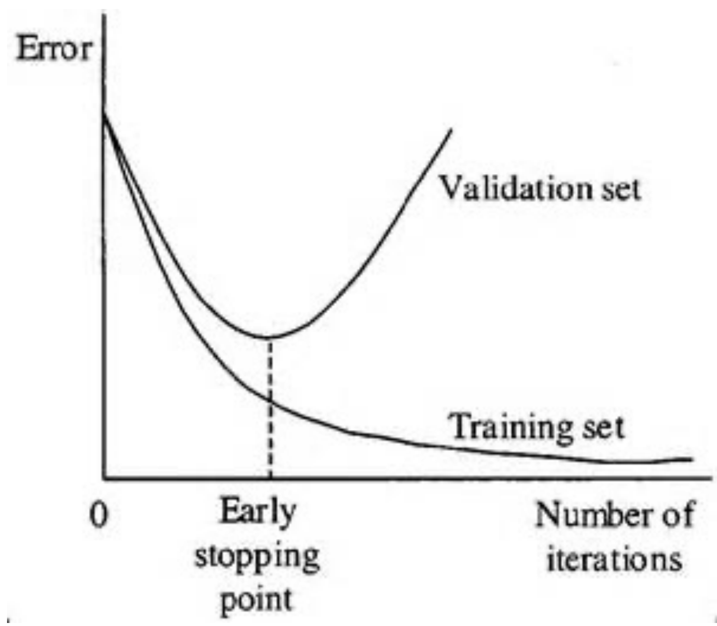


Schwessinger *et al.* bioRxiv 2019

Overfitting / Underfitting

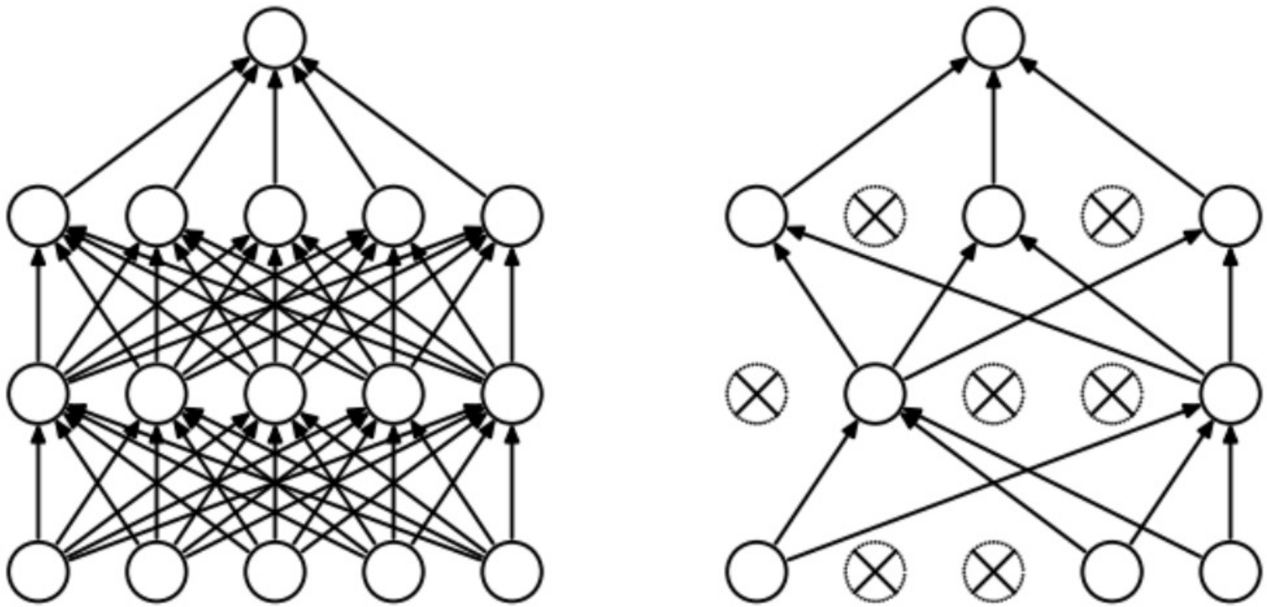
Tackle <u>overfitting</u> by in practice by:

- Optimizing regularization (L1 penalty, L2 penalty, Dropout)
- Use more data
- Early stopping
- Ensemble
- ...



Dropout

Dropout (mask) nodes <u>during training</u>, use all nodes for applying the network.



(a) Standard Neural Net (b) After applying dropout.

Overfitting / Underfitting

Tackle underfitting by in practice by:

- Increase the complexity / capacity of your model
- Use more data