

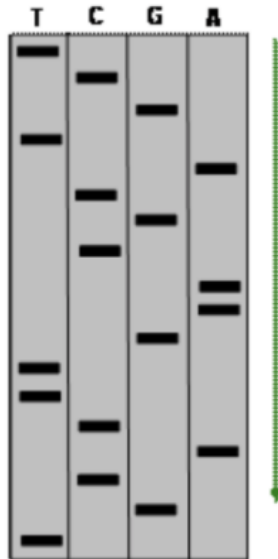
Deep Learning Project – DNA Sequencing

Part 1 : General presentation

Franco Alexis
Kenda Tom
Laloux Noé

LBRTI2101(B) – Data Science in Bioscience Engineering

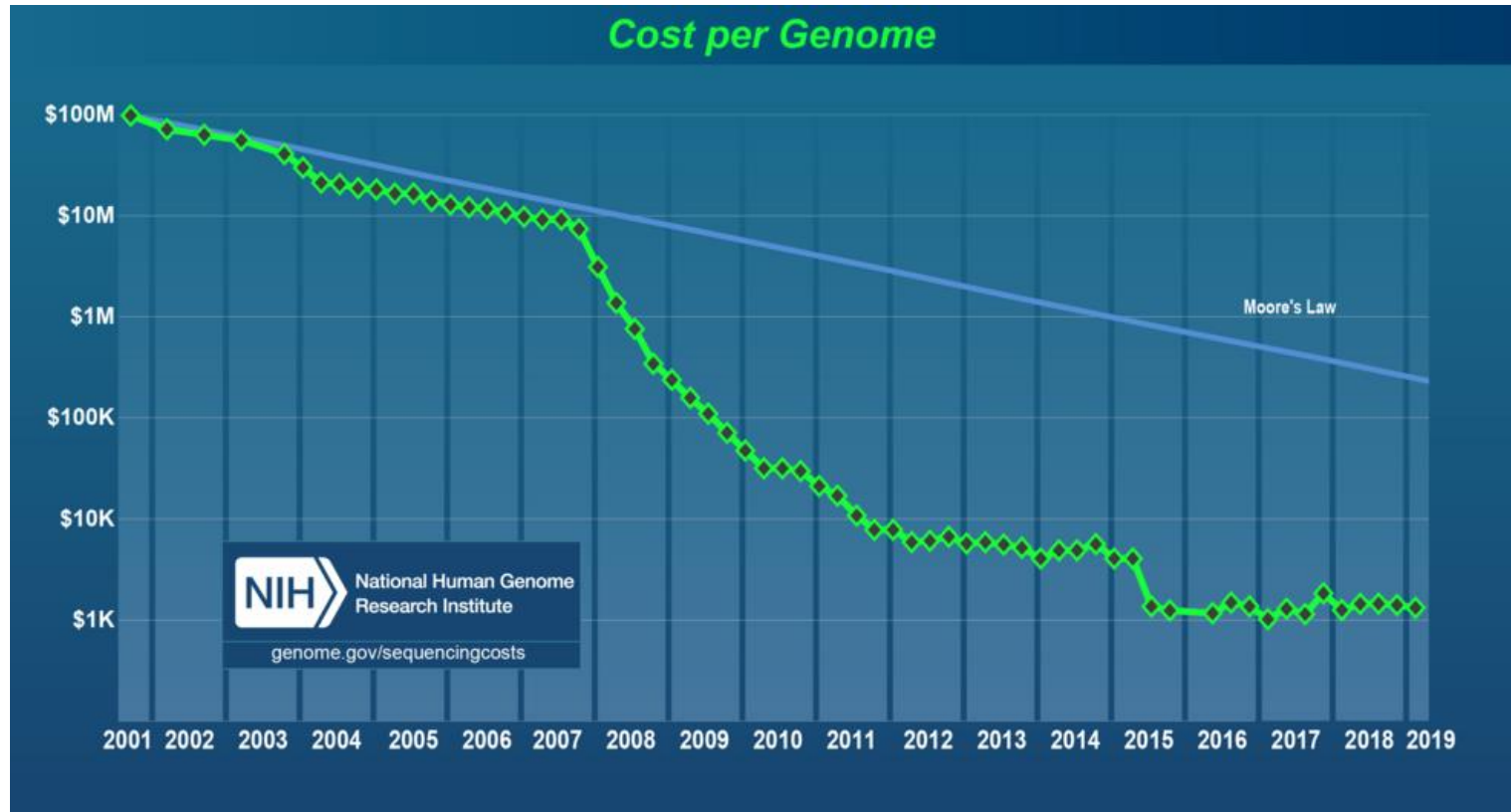
DNA Sequencing : 1980's vs Today



[Source](#)

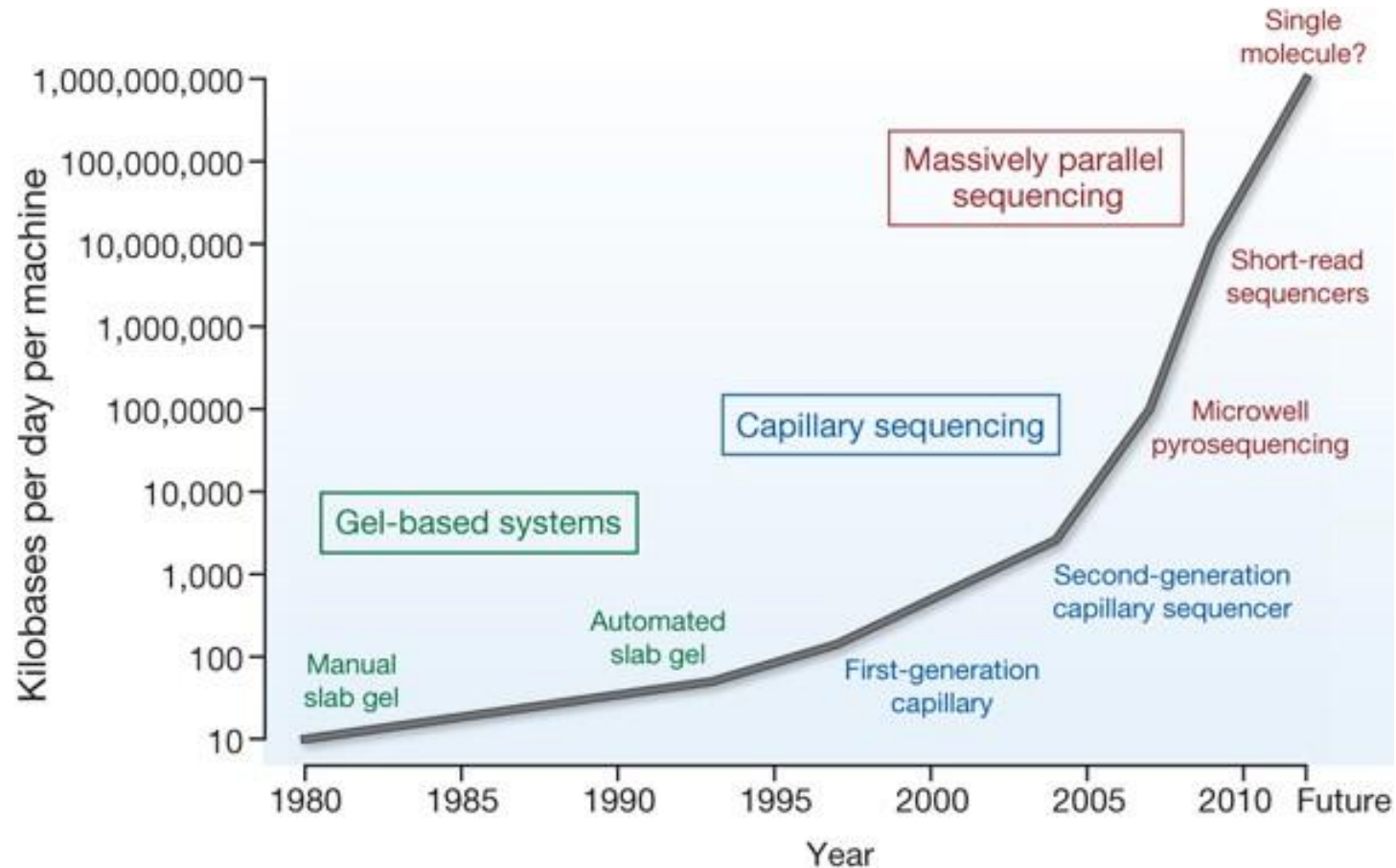


[Source](#)



[Source](#)

DNA Sequencing entering the Big Data scene



[Source](#)

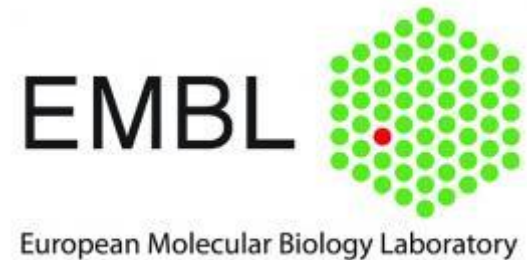
Computational tools demand to manage all this data



[Source](#)

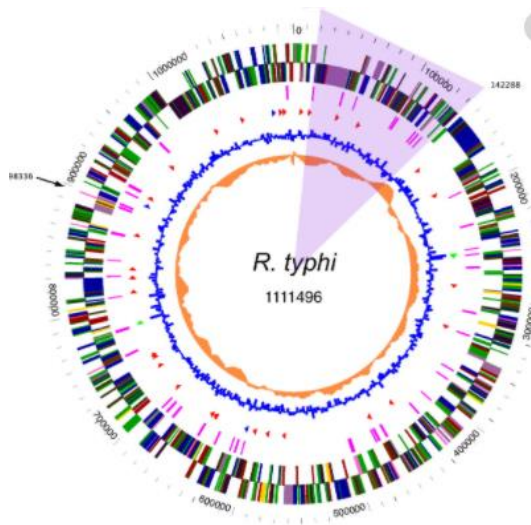


[Source](#)

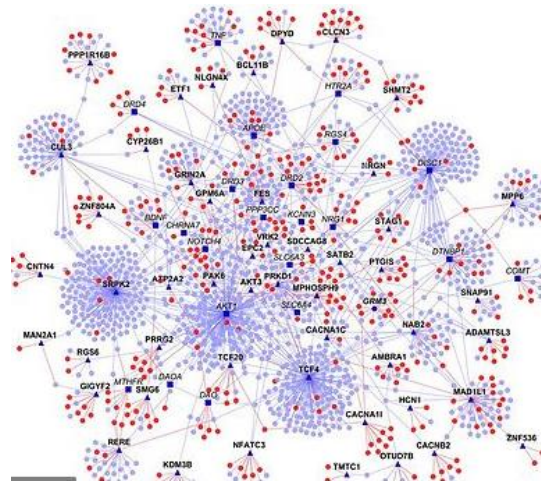


[Source](#)

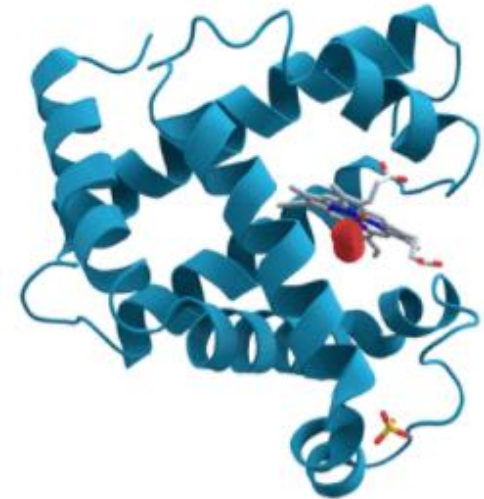
Applications of Bioinformatics Tools



[Source](#)



[Source](#)



[Source](#)

Real world applications

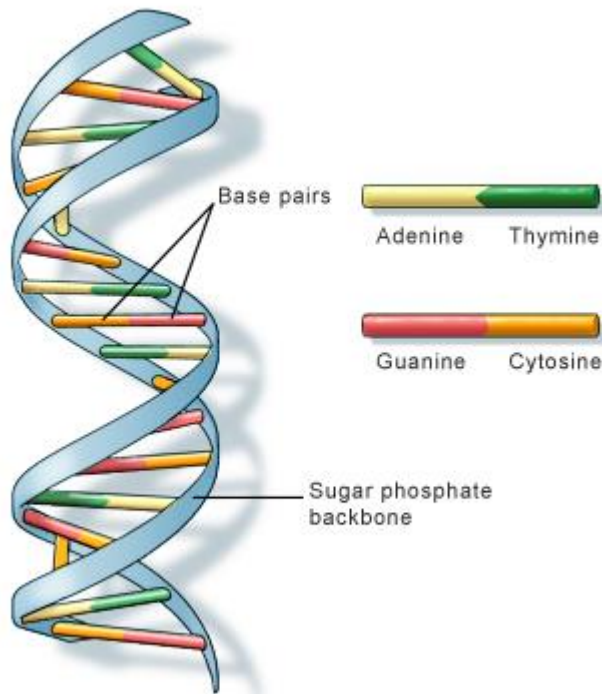


[Source](#)



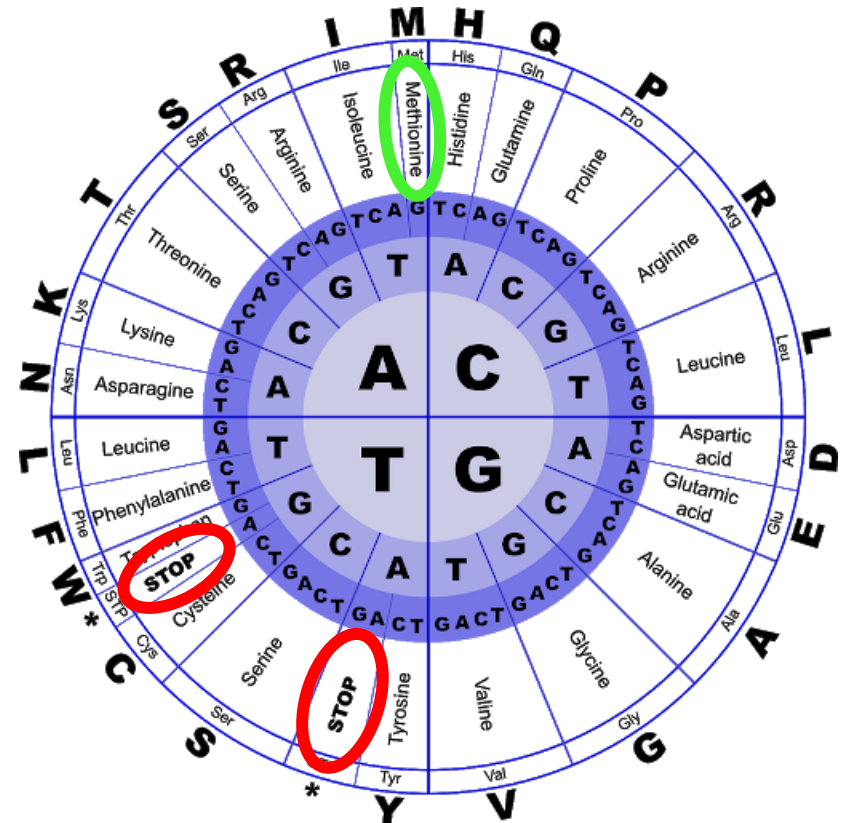
[Source](#)

DNA's structure and Protein synthesis



U.S. National Library of Medicine

[Source : News medical](#)



[Source : Fimalis](#)

What is the difference between ORFs and genes ?

RF : Reading Frame : 3 possibilities to read DNA for each strand

ex : (5'-P) TGGTATAGCAGAATGTCTGAATTTAGTTG (3'-OH)

-> TGG TAT AGC AGA **ATG** TCT GAA TTT AGT TG

-> T GGT ATA GCA GAA TGT CTG AAT TTA GTT G

-> TG GTA **TAG** CAG AAT GTC **TGA** ATT **TAG** TTG

ORF : Open Reading Frame : Part of a Reading Frame that starts with an ATG and ends with a STOP codon

ex : **ATG** TGG TAT AGC AGA ATG TCT GAA TTT AGT TGC **TGA**

Gene : ORF that corresponds to a real protein that can be expressed by the organism

ex : **ATG** TGG TAT AGC AGA ATG TCT GAA TTT AGT TGC **TGA**

-> **Met** - Trp - Tyr - Ser - Arg - Met - Ser - Glu - Phe - Ser - Cys -**STOP**

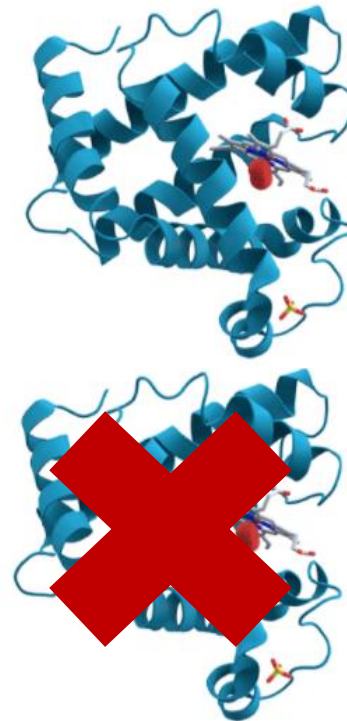
-> if this polypeptide corresponds to a protein

Goal of Our Project

Input :



Output :



How can Deep Learning help us ?

Our hypothesis :

- It is possible to predict whether or not a DNA sequence is a coding gene
- A well trained algorithm can do that through Deep Learning

Questions we have been facing in the beginning :

- Any existing libraries/algorithm doing this ?
- Is Deep Learning relevant for this application ?
- Can we find some data

2. Getting the coding sequences

Source : ncbi



Référence et bibliographie

[Index of the genome of saccharomyces cerevisiae](#)

[Genome data viewer \(with all the coding genes\) for S. cerevisiae](#)

LBIR1352 : Génétique Générale, Mahillon J.

LBPMC2201 : Bioinformatique : séquences d'ADN et de protéines, Ghislain M.

