"Genetics is where we come from. It's deeply natural to want to know."
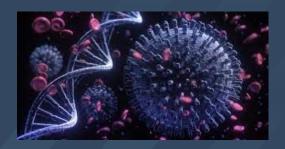
- Ellen Ullman

01

Objectives

DNA sequencing means determining the order of the four chemical building blocks, called "bases," that make up the DNA molecule.

In the DNA double helix, the four chemical bases always join with the same partner to form "base pairs." Adenine (A) always pairs with thymine (T); cytosine (C) always pairs with guanine (G).
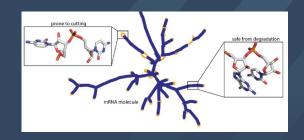
Designing a deep learning model that will predict a specific gen with a specific characteristic.

Background:

To design a stable mRNA vaccine that can survive shipment around the world as a single cut can render the entire vaccine useless.
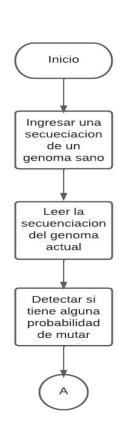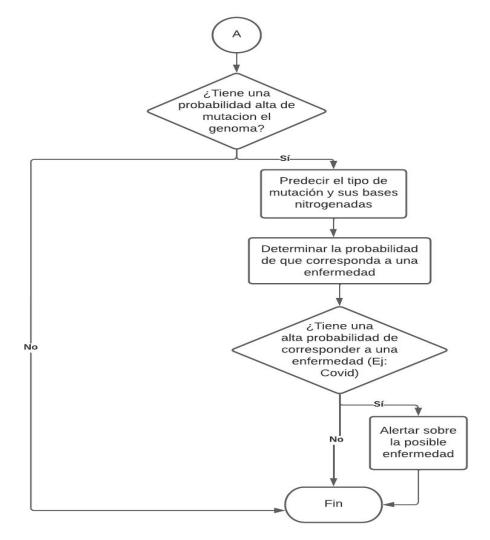

Alpha Fold is an AI system developed by DeepMind that predicts a protein's 3D structure from its amino acid sequence. It regularly achieves accuracy competitive with experiment.
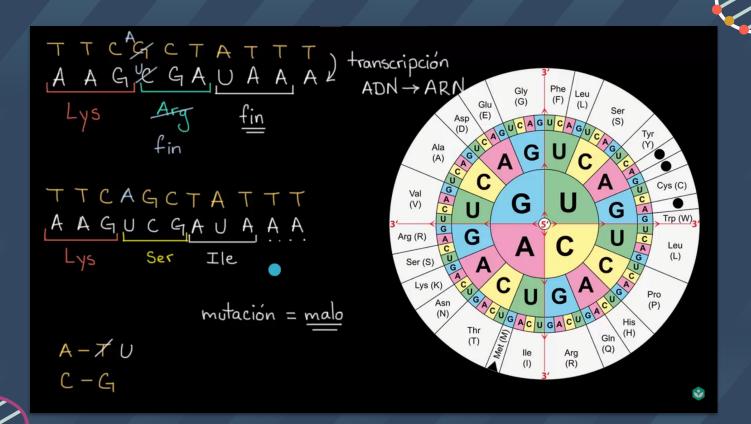
Proposed solution

A program that can predict the possibility of getting an illness related with a wrong sequencing of ARN

Inicio

Ingresar una secueciacion de un genoma sano

Leer la secuenciacion del genoma actual

Detectar si tiene alguna probabilidad de mutar

A

A

¿Tiene una probabilidad alta de mutacion el genoma?

No

Sí

Predecir el tipo de mutación y sus bases nitrogenadas

Determinar la probabilidad de que corresponda a una enfermedad

¿Tiene una alta probabilidad de corresponder a una enfermedad (Ej: Covid)

No

Sí

Alertar sobre la posible enfermedad

Fin

● A picture is worth a thousand words ●

02

Method

Markov Chains

They can model the transcription of nitrogenous bases

Conditional probability

$$\mathbb{P}(\delta_j | \delta_i) := \frac{\mathbb{P}(\delta_i \wedge \delta_j)}{\mathbb{P}(\delta_i)}$$

Update Rule

$$\Pi^{(t+1)} := \mathbb{A}\Pi^{(t)}$$

($\mathbb{A}$ stores initial probability distribution)

Implementation

Pandas

National Library of Medicine
National Center for Biotechnology Information

NLTK

@santiagoahl
@yeferssonaa

Thanks!

Do you have any questions?
sahumadal@unal.edu.co
yaldana@unal.edu.co