Cómo usar Python para curar enfermedades genéticas

Marina Moro López

pyladies MADRID

Python-Madrid

PyData Madrid

¡Hola! :D

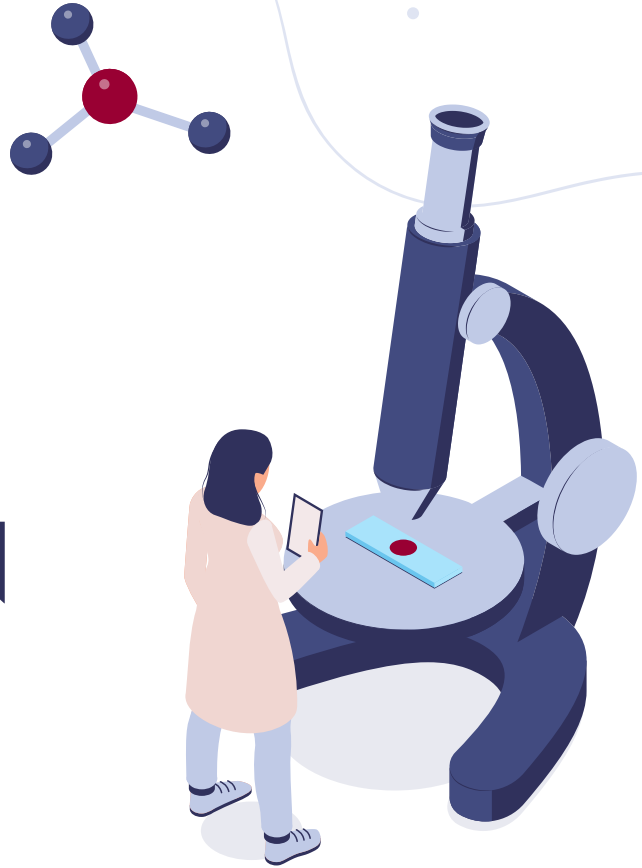
- Ingeniera biomédica
- Futura doctora en biofísica y bioingeniería
- ‘Programadora’ en mi día a día científico
- Secretaria de Python España

Índice

01

INTRODUCCIÓN

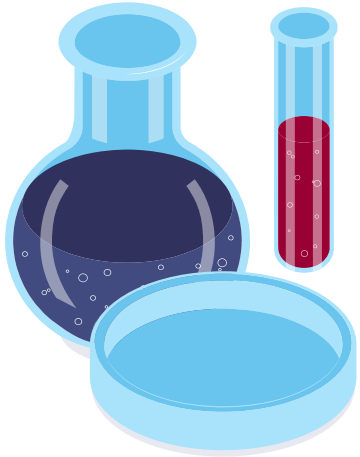
¿Qué es la epidermólisis bullosa?



¿Qué es la epidermólisis bullosa?

Enfermedad rara, genética y hereditaria que causa extrema fragilidad en la piel y mucosas

Producida por mutaciones en los genes de las proteínas de la piel



1 persona afectada cada 100.000 en España

Síntomas y tratamientos

Aparición de ampollas en la piel y mucosas por fricción o de forma espontánea

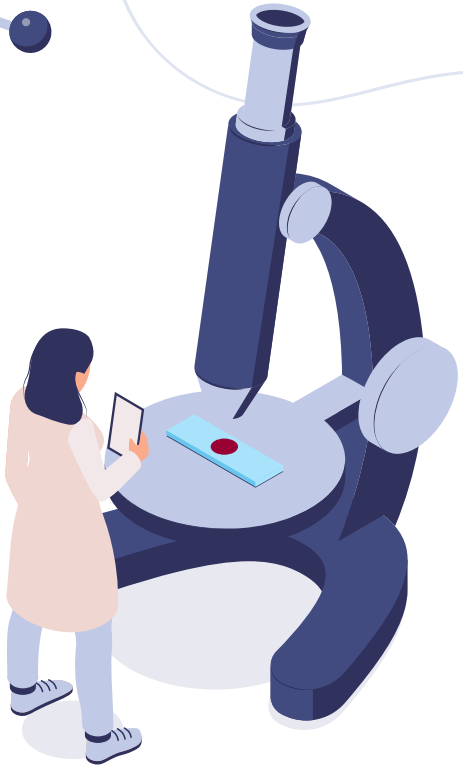
Complicaciones gastrointestinales, urinarias, pulmonares...
Cáncer
Limitación funcional y dependencia

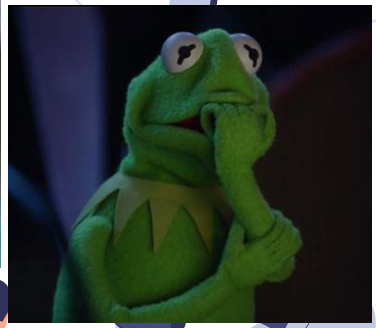
Tratamientos paliativos
Sin cura actualmente

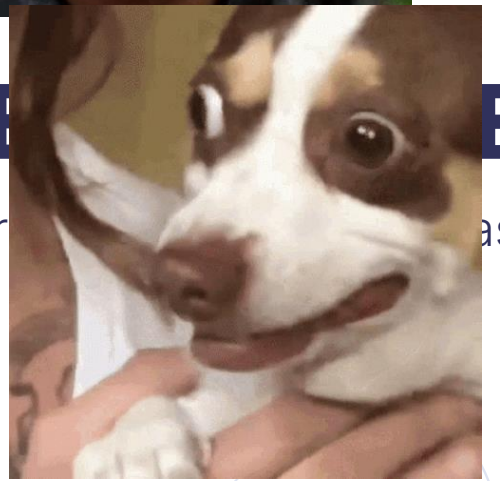
02

GENÉTICA BÁSICA

Teoría para entender el caso práctico







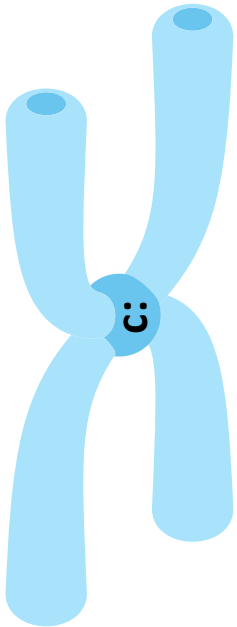
GE E

Teor as

Cromosoma - Gen - ADN

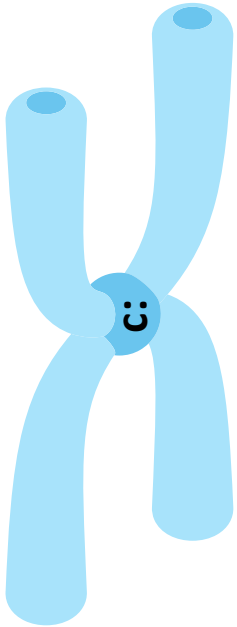
Cromosoma - Gen - ADN

Estructura que contiene
todos los genes

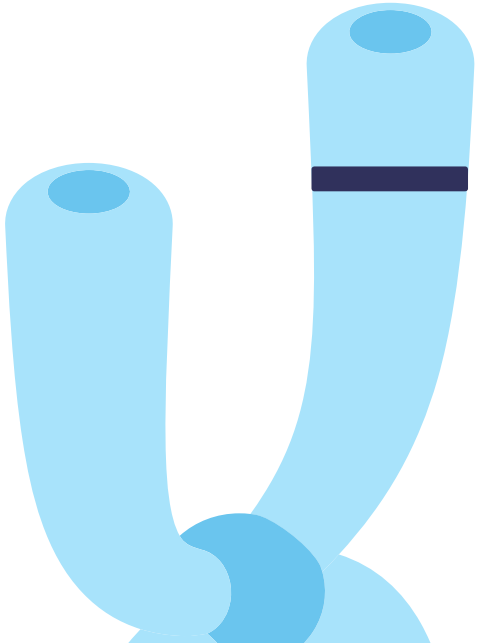


Cromosoma - Gen - ADN

Estructura que contiene todos los genes

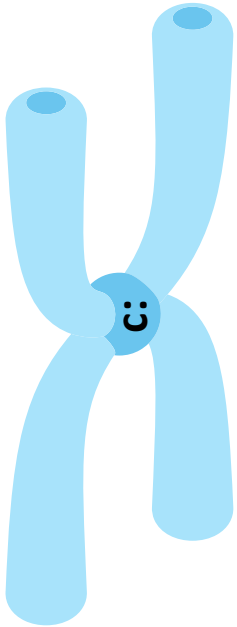


Segmento de ADN que determina un rasgo

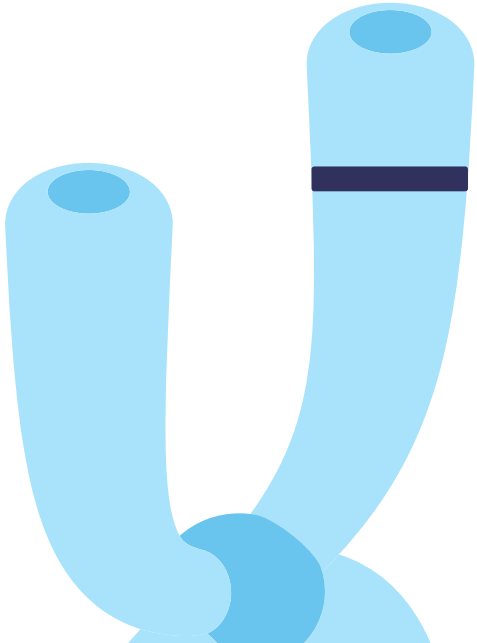


Cromosoma - Gen - ADN

Estructura que contiene todos los genes



Segmento de ADN que determina un rasgo

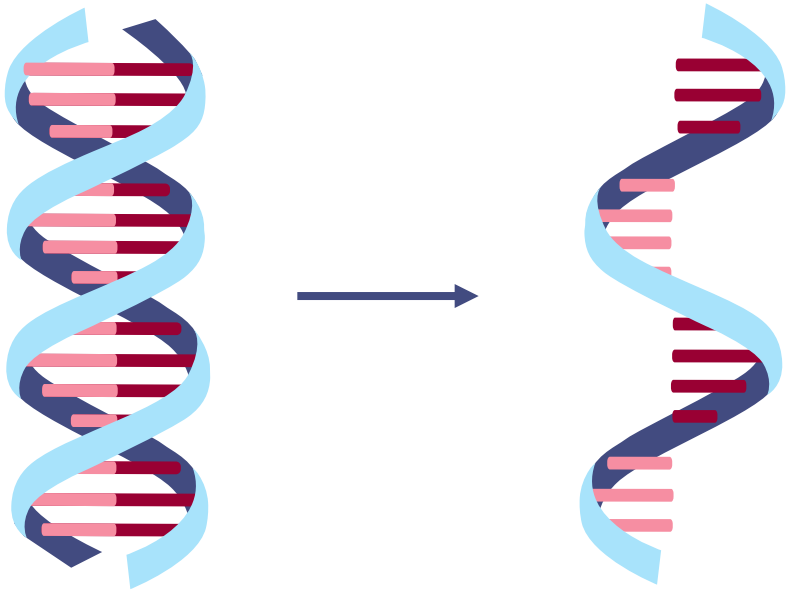


Doble hélice formada por bases

Bases del ADN

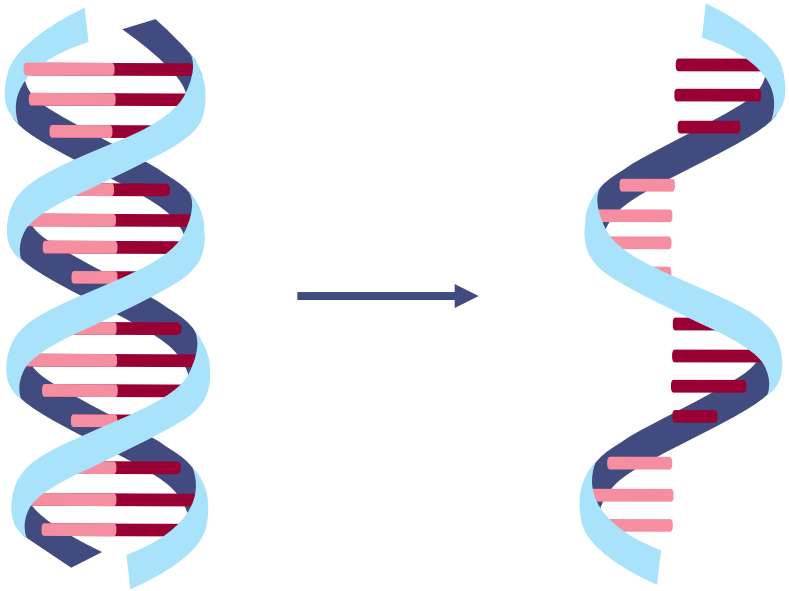
ADN - ARN - Proteína

ADN - ARN - Proteína



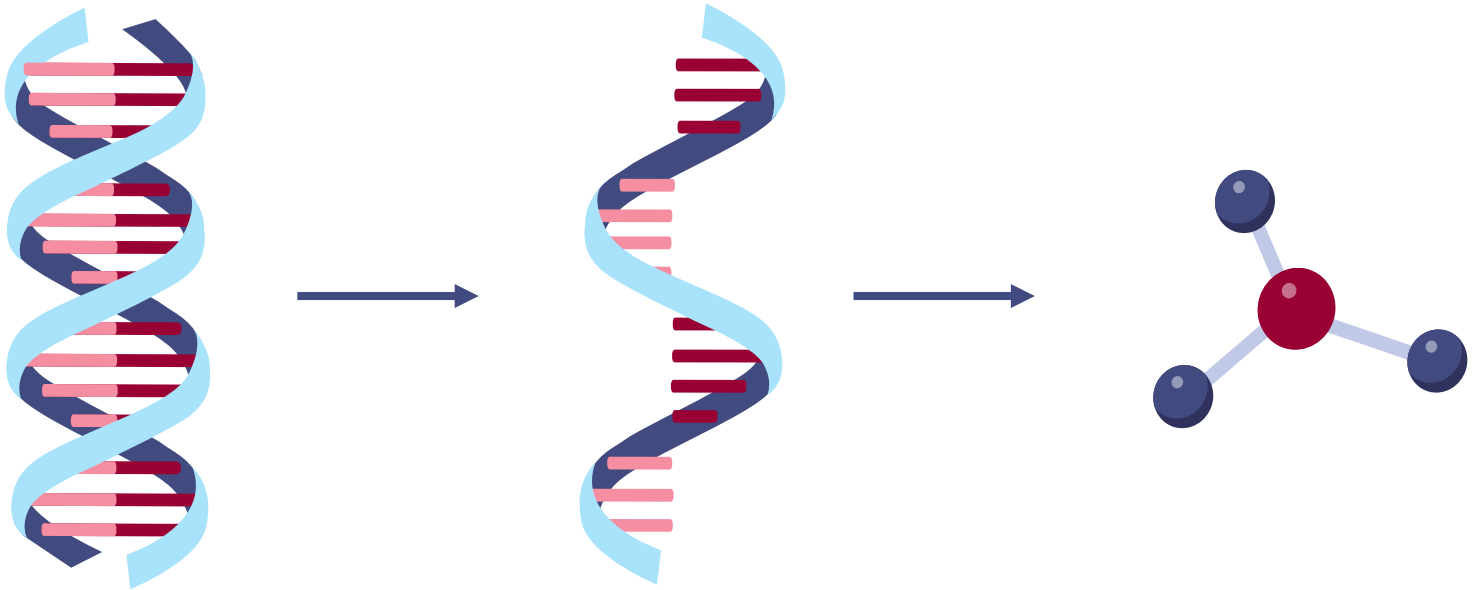
Hélice simple
T > U

ADN - ARN - Proteína



Hélice simple
T > U

ADN - ARN - Proteína

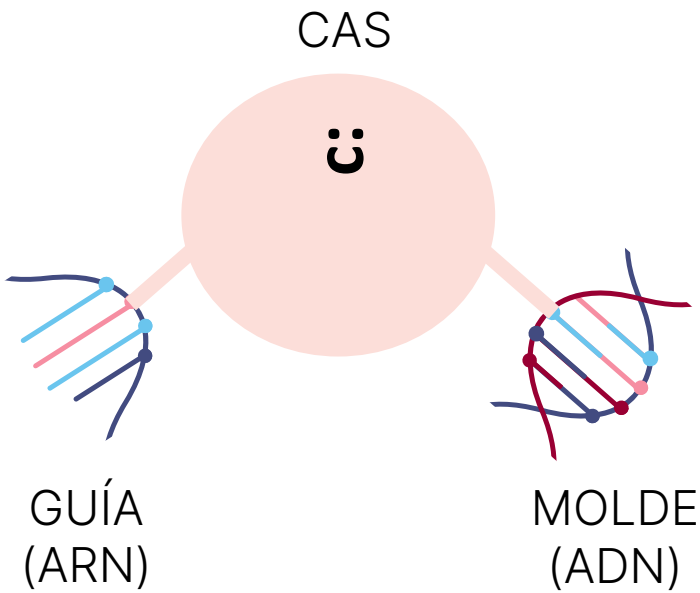


Hélice simple
T > U

CRISPR

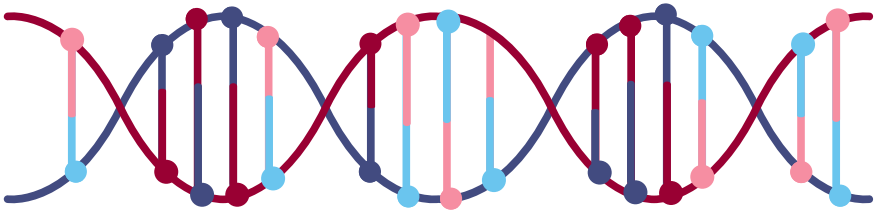
Corta y pega de secuencias de ADN (edición genética)

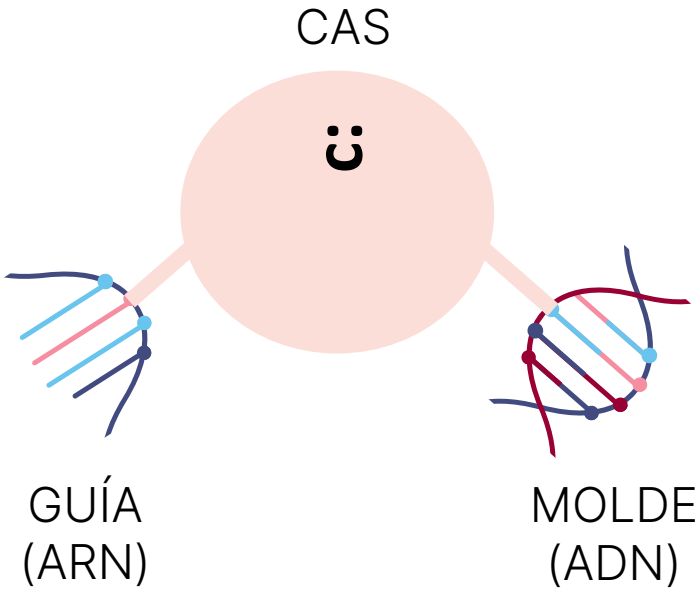
CRISPR



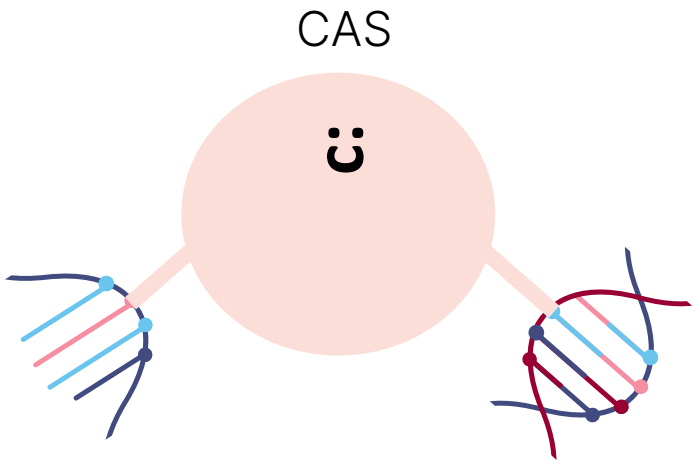
CRISPR

ADN
ORIGINAL

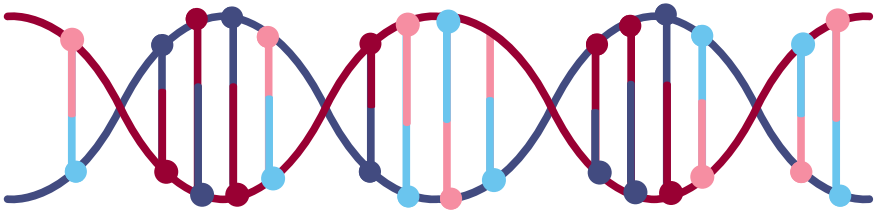




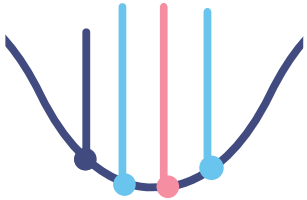
CRISPR



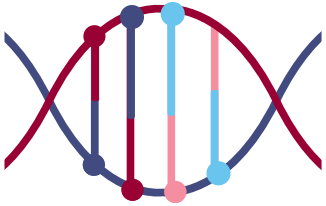
ADN
ORIGINAL



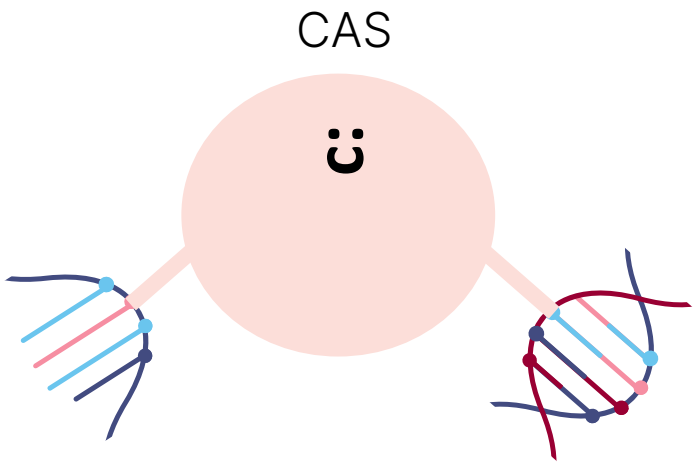
GUÍA
(ARN)

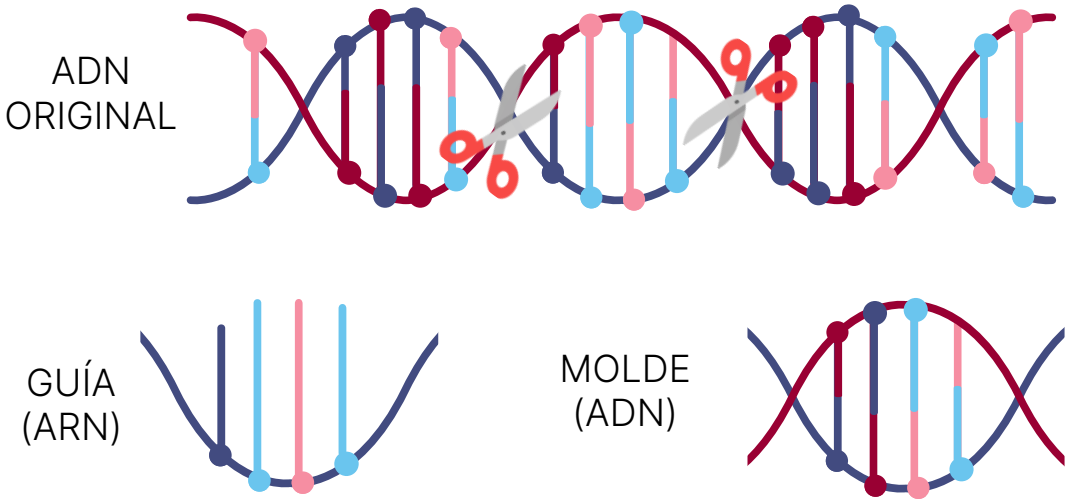


MOLDE
(ADN)

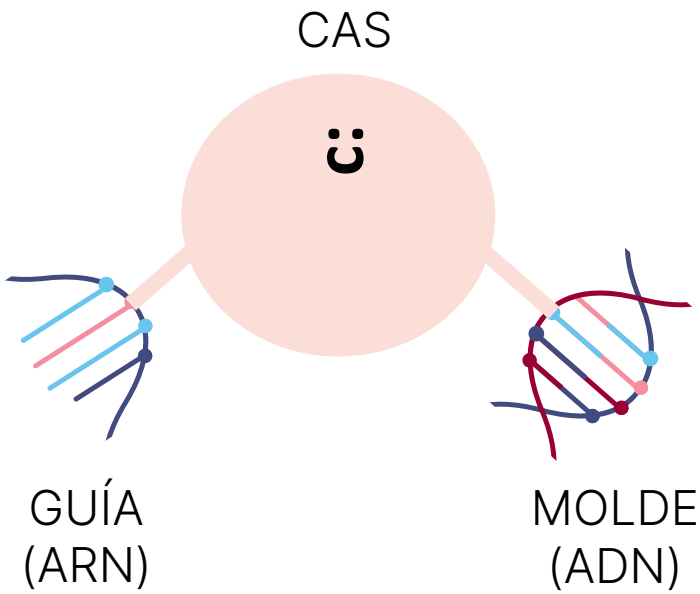


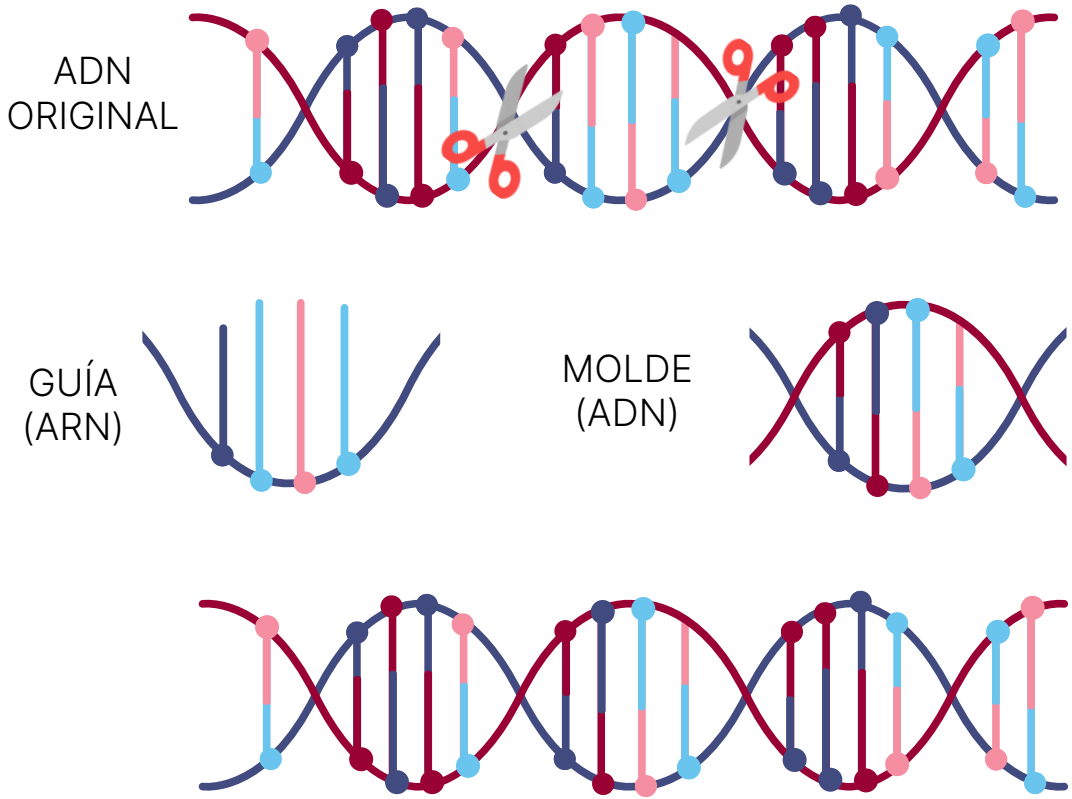
CRISPR





CRISPR

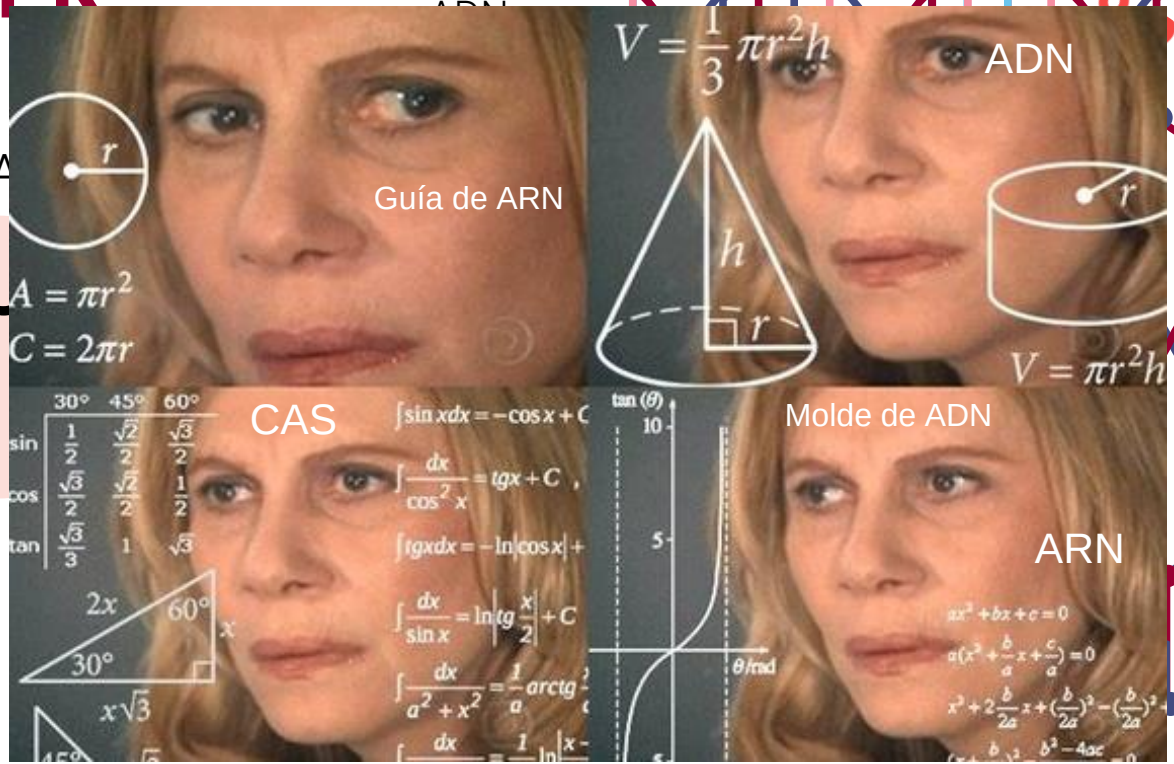




CRISPR

CA

¿

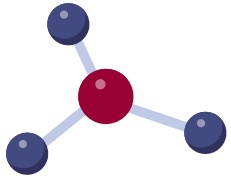


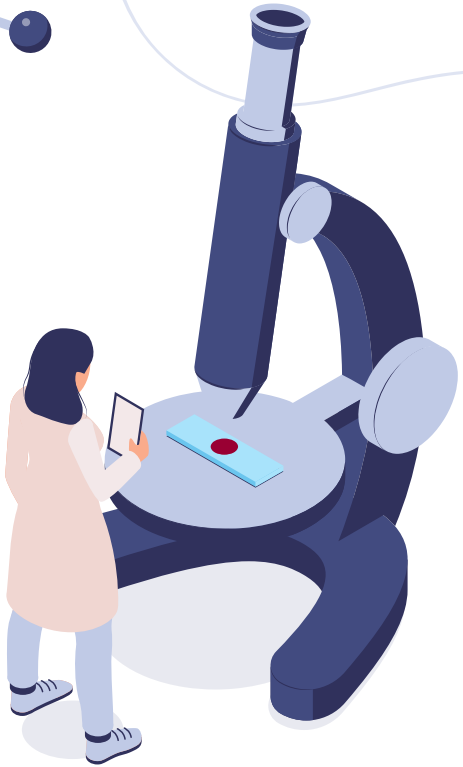
GUÍA
(ARN)

03

CASO PRÁCTICO

Tratamiento con CRISPR y Python





En el ADN de la paciente hay una C en vez de una T en la posición 6.527 del gen COL7A1

GEN COL7A1 DEFECTUOSO

…AAGAGATCCTGAGTCCCAGC CTGTTGCCACCCC…

Posición 6.527

En el ADN de la paciente hay una C en vez de una T en la posición 6.527 del gen COL7A1
Estrategia: borrar la C y poner una T en su lugar

GEN COL7A1 DEFECTUOSO

…AAGAGATCCTGAGTCCCAGC CTGTTGCCACCCC…

Posición 6.527

GEN COL7A1 CORREGIDO

…AAGAGATCCTGAGTCCTAGC CTGTTGCCACCCC…

marinamorolopez / **EB-PyMadDic2024**

```
from tkinter.filedialog import askopenfile

def main():

    gene_file = askopenfile(mode='r')
    gene_seq = gene_file.readlines()[1:]
    gene_seq = ''.join(gene_seq).replace('\n', '')

    DNA_guide, mutated_gene_seq, mold = knock_in_mid(gene_seq)
```

COL7A1 - Nucleotide - NCBI

ncbi.nlm.nih.gov/nuccore/?term=COL7A1

An official website of the United States government Here's how you know

National Library of Medicine
National Center for Biotechnology Information

Log in

Nucleotide

Nucleotide COL7A1 Search

Create alert Advanced Help

Species
Animals (2,456)
Bacteria (4)
Customize ...

Molecule types
genomic DNA/RNA (983)
mRNA (1,351)
Customize ...

Source databases
INSDC (GenBank) (661)
RefSeq (2,057)
Customize ...

Sequence Type
Nucleotide (2,713)
EST (4)
GSS (3)

Sequence length
Custom range...

Summary 20 per page Sort by Default order Send to:

GENE

Was this helpful?

COL7A1 – collagen type VII alpha 1 chain

Homo sapiens (human)

Also known as: EBD1, EBDCT, EBR1, NDNC8

Gene ID: 1294

Orthologs | Genome Data Viewer | RefSeq transcripts and proteins

RefSeq Sequences +

Filters: Manage Filters

Results by taxon

Top Organisms [Tree]
synthetic construct *(138)*
Homo sapiens *(60)*
Tyto alba *(20)*
Motacilla alba *(19)*
Prinia subflava *(16)*
All other taxa *(2467)*
More...

Find related data

Database: Select

Find items

Search details

7. **Homo sapiens** type VII collagen (**COL7A1**) mRNA, partial cds
1,875 bp linear mRNA
Accession: M65158.1 GI: 180914
Protein PubMed Taxonomy
GenBank FASTA Graphics

8. Human alpha-1 type VII collagen (**COL7A1**) mRNA, complete cds
9,287 bp linear mRNA
Accession: L02870.1 GI: 987124
Protein PubMed Taxonomy
GenBank FASTA Graphics

9. **Homo sapiens** collagen type VII alpha 1 chain (**COL7A1**), RefSeqGene (LRG_286) on chromosome 3; nuclear gene for mitochondrial product
38,180 bp linear DNA
Accession: NG_007065.1 GI: 160358345
Protein PubMed Taxonomy
GenBank FASTA Graphics

10. PREDICTED: **Homo sapiens** collagen type VII alpha 1 chain (**COL7A1**), transcript variant X17, mRNA
6,349 bp linear mRNA
Accession: XM_054345179.1 GI: 2462587008
BioProject Protein Taxonomy
GenBank FASTA Graphics

11. PREDICTED: **Homo sapiens** collagen type VII alpha 1 chain (**COL7A1**), transcript variant X16, mRNA

National Library of Medicine
National Center for Biotechnology Information

Log in

Nucleotide | Nucleotide | Search | Advanced | Help

GenBank

Send to:

Complete Record
Coding Sequences
Gene Features

Choose Destination

File | Clipboard | Collections | Analysis Tool

Download 1 item.

Format: FASTA

Show GI

Create File

Homo sapiens collagen type VII alpha 1 chain (COL7A1), RefSeqG on chromosome 3; nuclear gene for mitochondrial product

NCBI Reference Sequence: NG_007065.1

FASTA Graphics

Go to:

```
LOCUS       NG_007065              38180 bp    DNA     linear   PRI 10-OCT-2024
DEFINITION  Homo sapiens collagen type VII alpha 1 chain (COL7A1), RefSeqGene
            (LRG_286) on chromosome 3; nuclear gene for mitochondrial product.
ACCESSION   NG_007065
VERSION     NG_007065.1
KEYWORDS    RefSeq; RefSeqGene.
SOURCE      Homo sapiens (human)
  ORGANISM  Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Euarchontoglires; Primates; Haplorrhini;
            Catarrhini; Hominidae; Homo.
REFERENCE   1  (bases 1 to 38180)
```

Change region shown

PubMed

Taxonomy

```
DNA_guide, mutated_gene_seq, mold = knock_in_mid(gene_seq)
```

```
mutated_gene_file = open('MUTATED_SEQUENCE.txt', 'w')
mutated_gene_file.write(mutated_gene_seq)
mutated_gene_file.close()

guide_file = open('GUIDE.txt', 'w')
guide_file.write(DNA_to_RNA(DNA_guide))
guide_file.close()

mold_file = open('MOLD.txt', 'w')
mold_file.write(mold)
mold_file.close()
```

```
DNA_guide, mutated_gene_seq, mold = knock_in_mid(gene_seq)
```

```
mutated_gene_file = open('MUTATED_SEQUENCE.txt', 'w')
mutated_gene_file.write(mutated_gene_seq)
mutated_gene_file.close()

guide_file = open('GUIDE.txt', 'w')
guide_file.write(DNA_to_RNA(DNA_guide))
guide_file.close()

mold_file = open('MOLD.txt', 'w')
mold_file.write(mold)
mold_file.close()
```

```
def DNA_to_RNA(DNA_guide):

    RNA_guide = ""
    for base in DNA_guide:
        if base == "T":
            RNA_guide += "A"
        elif base == "A":
            RNA_guide += "U"
        elif base == "C":
            RNA_guide += "G"
        elif base == "G":
            RNA_guide += "C"

    return RNA_guide
```

```python
def knock_in_mid(gene_seq):

    mutation_position = int(input("Introduce the numeric position of the mutation base (e.g. 1, 25, 203): "))
    while mutation_position <= 0:
        print('Invalid input. Introduce positive number. ')
        mutation_position = int(input("Introduce the numeric position of the mutation base (e.g. 1, 25, 203): "))

    mutation_base = input("Introduce the new base(s) to be in the defined mutation position (e.g. A, CAG, CCGT, GACTA): ")

    DNA_guide = gene_seq[mutation_position-25:mutation_position+25]
    mold = DNA_guide[:]
    mutated_gene_seq = gene_seq[:mutation_position-1] + mutation_base + gene_seq[mutation_position:]
    mold = mutated_gene_seq[mutation_position-25:mutation_position+25]

    return DNA_guide, mutated_gene_seq, mold
```

En el ADN de la paciente hay una C en vez de una T en la posición 6.527 del gen COL7A1

Estrategia: borrar la C y poner una T en su lugar

ADN A CORREGIR

CAAAGGAGCCCCTGGGGAGTTTGTTAACCTGGGGGTTCCAGAGGCAAGTGGGTGTGGAGAATGACAGAAC
GAAGGGACCCCTCAAGAGAGCCTGATACCCGTAACCCTCACCCTAGAGGCCTCCTCAAGGCCAGGGCCAG
AAGAGATCCTGAGTCCCAGCCTGTTGCCACCCCTACCCTCCAACAGGACAGAGTTCGGCCTGGATGCACT
TGGCTCTGGGGGTGATGTGATCCGCGCCATCCGTGAGCTTAGCTACAAGGGGGGCAACACTCGCACAGGG
GCTGCAATTCTCCATGTGGCTGACCATGTCTTCCTGCCCCAGCTGGCCCGACCTGGTGTCCCCAAGGTGA

GUÍA (ARN)

UCCCGGUCUUCUCUAGGACUCAGGGUCGGACAACGGUGGGGAUGGGAGGU

En el ADN de la paciente hay una C en vez de una T en la posición 6.527 del gen COL7A1
Estrategia: borrar la C y poner una T en su lugar

ADN A CORREGIR

CAAAGGAGCCCCTGGGGAGTTTGTTAACCTGGGGGTTCCAGAGGCAAGTGGGTGTGGAGAATGACAGAAC
GAAGGGACCCCTCAAGAGAGCCTGATACCCGTAACCCTCACCCTAGAGGCCTCCTCAAGGCCAGGGCCAG
AAGAGATCCTGAGTCCCAGCCTGTTGCCACCCCTACCCTCCAACAGGACAGAGTTCGGCCTGGATGCACT
TGGCTCTGGGGGTGATGTGATCCGCGCCATCCGTGAGCTTAGCTACAAGGGGGGCAACACTCGCACAGGG
GCTGCAATTCTCCATGTGGCTGACCATGTCTTCCTGCCCCAGCTGGCCCGACCTGGTGTCCCCAAGGTGA

MOLDE (ADN)

AGGGCCAGAAGAGATCCTGAGTCCTAGCCTGTTGCCACCCCTACCCTCCA

En el ADN de la paciente hay una C en vez de una T en la posición 6.527 del gen COL7A1

Estrategia: borrar la C y poner una T en su lugar

ADN A CORREGIR

CAAAGGAGCCCCTGGGGAGTTTGTTAACCTGGGGGTTCCAGAGGCAAGTGGGTGTGGAGAATGACAGAAC
GAAGGGACCCCTCAAGAGAGCCTGATACCCGTAACCCTCACCCTAGAGGCCTCCTCAAGGCCAGGGCCAG
AAGAGATCCTGAGTCCCAGCCTGTTGCCACCCCTACCCTCCAACAGGACAGAGTTCGGCCTGGATGCACT
TGGCTCTGGGGGTGATGTGATCCGCGCCATCCGTGAGCTTAGCTACAAGGGGGGCAACACTCGCACAGGG
GCTGCAATTCTCCATGTGGCTGACCATGTCTTCCTGCCCCAGCTGGCCCGACCTGGTGTCCCCAAGGTGA

ADN CORREGIDO

CAAAGGAGCCCCTGGGGAGTTTGTTAACCTGGGGGTTCCAGAGGCAAGTGGGTGTGGAGAATGACAGAAC
GAAGGGACCCCTCAAGAGAGCCTGATACCCGTAACCCTCACCCTAGAGGCCTCCTCAAGGCCAGGGCCAG
AAGAGATCCTGAGTCCTAGCCTGTTGCCACCCCTACCCTCCAACAGGACAGAGTTCGGCCTGGATGCACT
TGGCTCTGGGGGTGATGTGATCCGCGCCATCCGTGAGCTTAGCTACAAGGGGGGCAACACTCGCACAGGG
GCTGCAATTCTCCATGTGGCTGACCATGTCTTCCTGCCCCAGCTGGCCCGACCTGGTGTCCCCAAGGTGA

```
def knock_in_end(gene_seq):

    plasmid_file = askopenfile(mode='r')
    plasmid_seq = plasmid_file.readlines()[1:]
    plasmid_seq = ''.join(plasmid_seq).replace('\n', '')

    DNA_guide = gene_seq[len(gene_seq)-50:len(gene_seq)]
    mutated_gene_seq = gene_seq + plasmid_seq
    mold = DNA_guide + plasmid_seq

    return DNA_guide, mutated_gene_seq, mold
```

```
def knock_out(gene_seq):

    DNA_guide = gene_seq
    mutated_gene_seq = ""
    mold = ""

    return DNA_guide, mutated_gene_seq, mold
```

```python
def repeated_seq(gene_seq):

    mutation_position = int(input("Introduce the numeric position of the mutation base (e.g. 1, 25, 203): "))
    while mutation_position <= 0:
        print('Invalid input. Introduce positive number. ')
        mutation_position = int(input("Introduce the numeric position of the mutation base (e.g. 1, 25, 203): "))

    rep_letters = input("Introduce the letters that are repeated (e.g. AAT, CAG, CCGT, GACTA): ")

    rep_number = int(input("Introduce the healthy number of repetitions (e.g. 20, 35, 42): "))
    while rep_number <= 0:
        print('Invalid input. Introduce positive number. ')
        rep_number = int(input("Introduce the healthy number of repetitions (e.g. 20, 35, 42): "))

    search_repeat = gene_seq.find(rep_letters)
    DNA_guide = rep_letters * len(search_repeat)
    mold = rep_letters * rep_number
    mutated_gene_seq = gene_seq[:mutation_position-1] + mold + gene_seq[mutation_position+len(search_repeat)*len(rep_letters):]

    return DNA_guide, mutated_gene_seq, mold
```

Timeline del tratamiento

01

Obtención de datos

Secuenciación genética de la paciente

02

Python

Automatiza el diseño de CRISPR

03

Síntesis

Producción bioquímica del sistema con guía y molde

04

Inyección

In vivo / ex vivo

Más aplicaciones terapéuticas

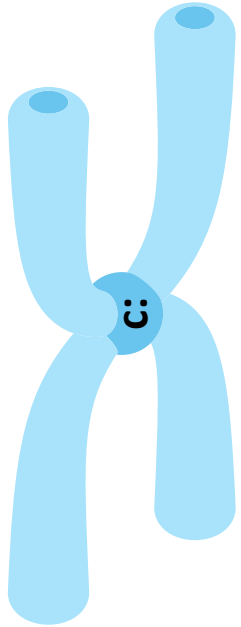
Células CAR-T contra el cáncer

Terapia antiviral (SARS-CoV-2, VIH)

Lucha contra enfermedades infecciosas (malaria, fiebre amarilla)

Bibliografía de interés

- C. Sánchez-Jimeno et al., “Diagnóstico genético de la epidermólisis bullosa: recomendaciones de un grupo español de expertos”, Actas Dermo, vol. 109(2), pp. 104-122, Mar 2018.
- M. García et al., “Preclinical model for phenotypic correction of dystrophic epidermolysis bullosa by in vivo CRISPR-Cas9 delivery using adenoviral vectors”, Molecular Therapy, vol. 27, pp. 96-108, Dec 2022.
- J. Bonafont et al., “Clinically Relevant Correction of Recessive Dystrophic Epidermolysis Bullosa by Dual sgRNA CRISPR/Cas9-Mediated Gene Editing”, Molecular Therapy, vol. 27(5), pp. 986-998, May 2019.
- J. Bischof et al., “Emerging Gene Therapeutics for Epidermolysis Bullosa under Development”, Molecular Sciences, vol. 25(4), pp. 2243, Feb 2024.
- J. P. Hofbauer et al., “Challenges and progress related to gene editing in rare skin diseases”, Drug Delivery Reviews, vol. 208, May 2024.



¡Muchas gracias!

¿Preguntas?

@marinamorolopez

Marina Moro López

marinamorolopez

EB-PyMadDic2024 Public

pyladies MADRID

Python-Madrid

PyData Madrid


Dijiste que era un
meetup de Python


De Python Y genética

¡Muchas gracias!

¿Preguntas?

@marinamorolopez

Marina Moro López

marinamorolopez

EB-PyMadDic2024 Public

pyladies MADRID

Python-Madrid

PyData Madrid