

Diagnóstico prenatal de enfermedades genéticas usando Python

Marina Moro López



Women
Techmakers



whoami

- Ingeniera biomédica
- Futura doctora en biomedicina
- 'Programadora' en el ámbito científico



Índice

1

¿Qué es el
diagnóstico
prenatal?

2

Conceptos básicos
de genética

3

Caso práctico
de diagnóstico
prenatal

4

Más aplicaciones
clínicas con
genética y Python

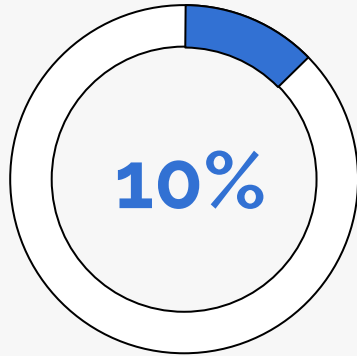


1

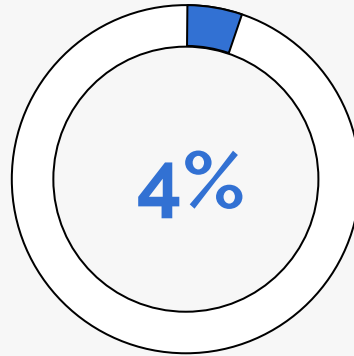
**¿Qué es el
diagnóstico
prenatal?**

Contexto

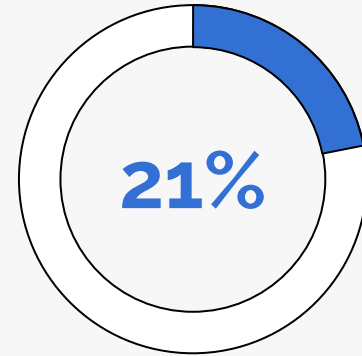
Existen alrededor de 6.000 enfermedades genéticas



Población que sufre una enfermedad genética a lo largo de su vida



Neonatos con una enfermedad genética



Muertes de neonatos debido a enfermedades genéticas

Diagnóstico prenatal

Identificación de enfermedades genéticas antes del nacimiento

Por edad o mutaciones de los progenitores, enfermedades genéticas en hijxs existentes o indicios en los resultados de otras pruebas clínicas

Diagnóstico prenatal

Identificación de enfermedades genéticas antes del nacimiento

Por edad o mutaciones de los progenitores, enfermedades genéticas en hijxs existentes o indicios en los resultados de otras pruebas clínicas

Ultrasonido

Defectos a simple vista

Análisis hormonal

Hormonas en la sangre de la persona gestante

ADN fetal

Mutaciones del ADN

NO INVASIVO

Diagnóstico prenatal

Identificación de enfermedades genéticas antes del nacimiento

Por edad o mutaciones de los progenitores, enfermedades genéticas en hijxs existentes o indicios en los resultados de otras pruebas clínicas

Ultrasonido

Defectos a simple vista

Análisis hormonal

Hormonas en la sangre de la persona gestante

ADN fetal

Mutaciones del ADN

NO INVASIVO

INVASIVO

Amniocentesis

Defectos en cromosomas

Vellosidades coriónicas

Defectos en cromosomas y mutaciones del ADN

Biopsia de placenta

Defectos en cromosomas

Diagnóstico prenatal

Identificación de enfermedades genéticas antes del nacimiento

Por edad o mutaciones de los progenitores, enfermedades genéticas en hijos existentes o indicios en los resultados de otras pruebas clínicas

Ultrasonido

Defectos a simple vista

Análisis hormonal

Hormonas en la sangre de la persona gestante

ADN fetal

Mutaciones del ADN

NO INVASIVO

INVASIVO

Amniocentesis

Defectos en cromosomas

Vellosidades coriónicas

Defectos en cromosomas y mutaciones del ADN

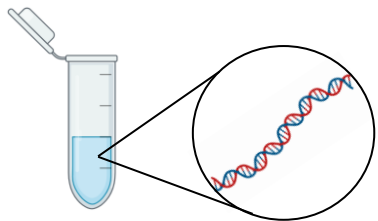
Biopsia de placenta

Defectos en cromosomas

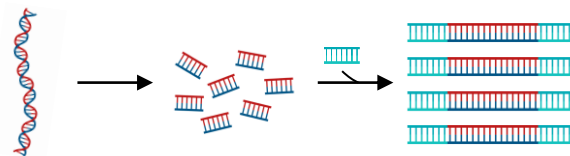
* **Diagnóstico preimplantacional**: análisis del material genético del embrión antes de implantarlo *in vitro*

Workflow

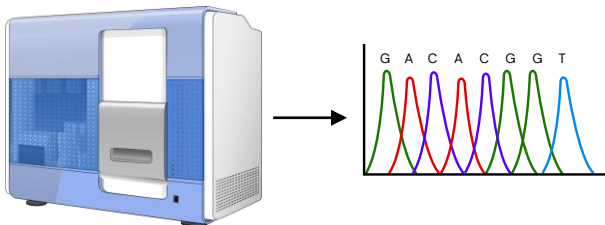
1. Extracción del ADN



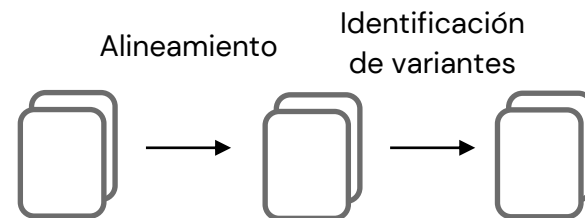
2. Preparación del ADN



3. Secuenciación



4. Análisis





2

Conceptos básicos de genética

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

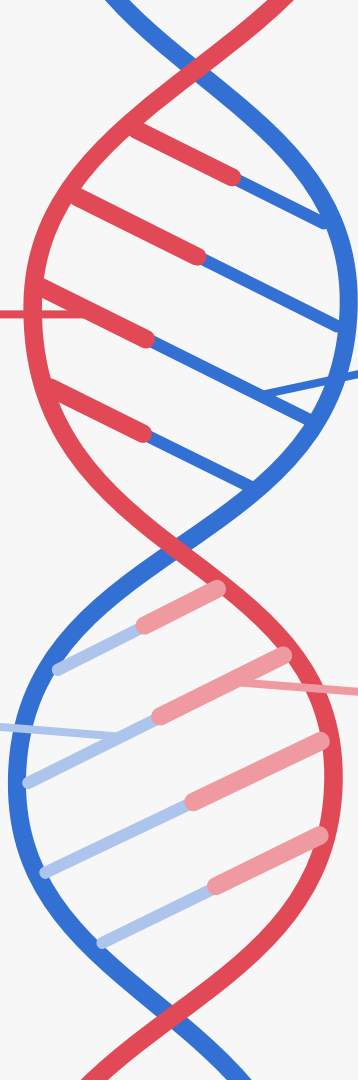


ADENINA (A)

TIMINA (T)

CITOSINA (C)

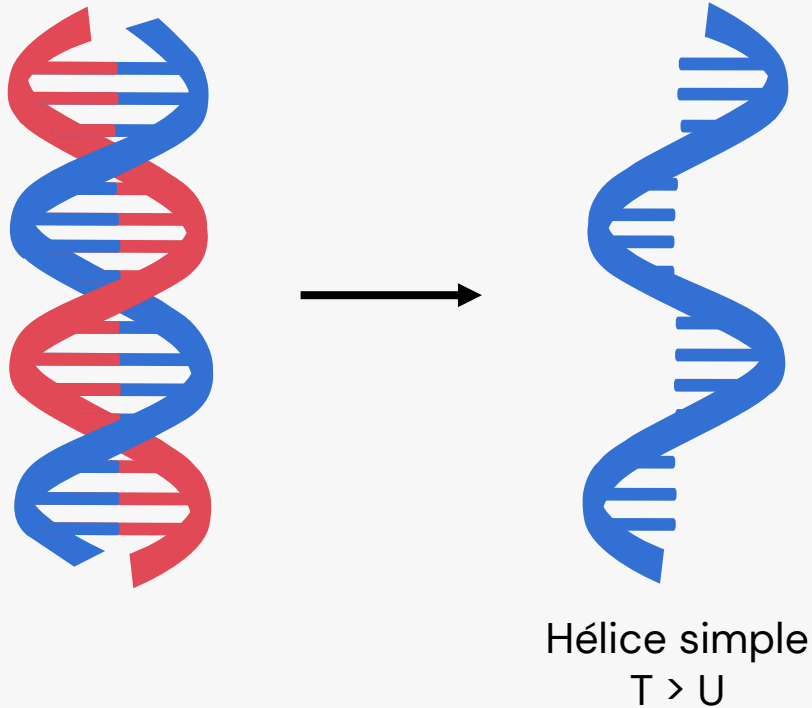
GUANINA (G)



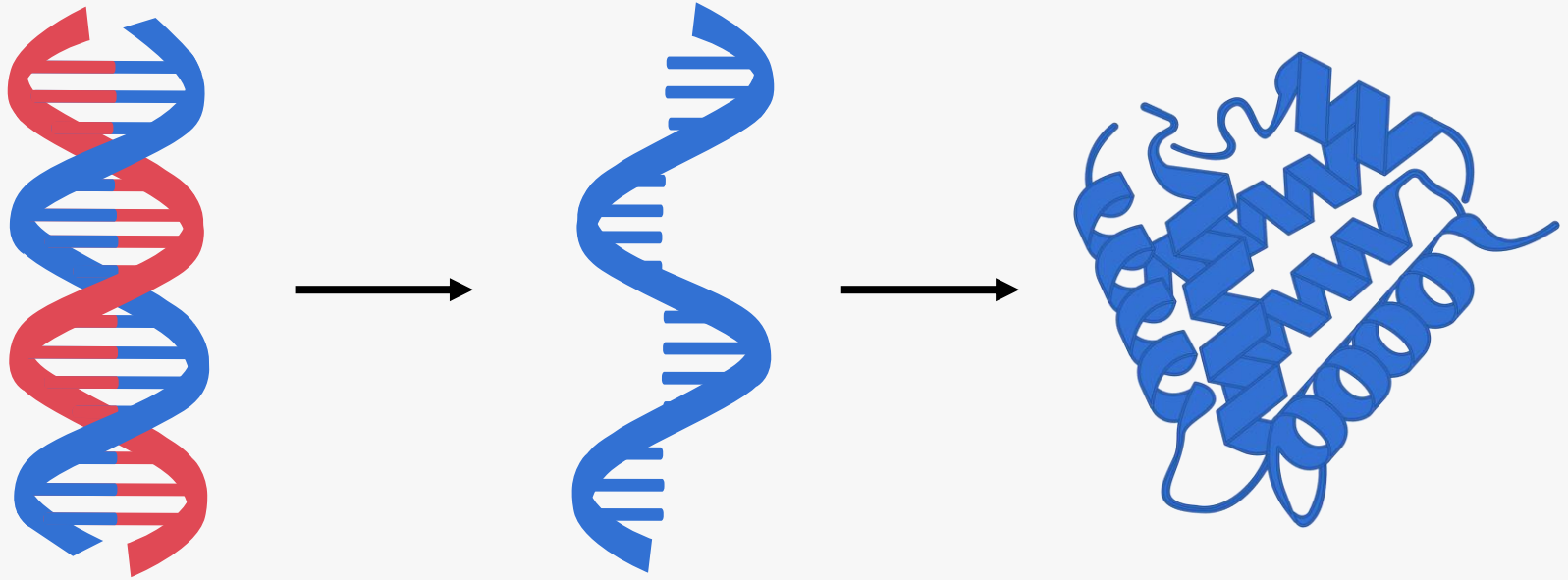
ADN - ARN - proteína



ADN - ARN - proteína



ADN - ARN - proteína



Hélice simple
 $T > U$



3

Caso práctico de diagnóstico prenatal

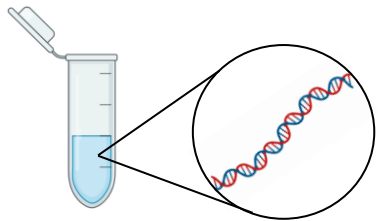
Epidermólisis bullosa

- Enfermedad genética rara (1 de cada 60.000 recién nacidos) que afecta a las mucosas y la piel
- Provoca fragilidad extrema en mucosas y piel, apareciendo ampollas con el contacto o incluso de forma espontánea (piel de mariposa)
- Otras complicaciones son lesiones dentales, atrofia muscular, estrechamiento de las vías respiratorias, gastrointestinales y urogenitales, y cáncer

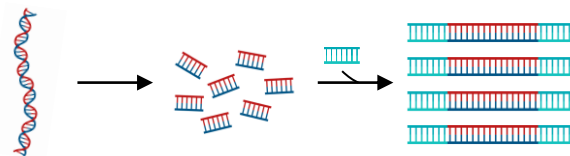
Causada por mutaciones en genes de proteínas de la piel, como colágeno VII, lamininas, integrinas...

Workflow

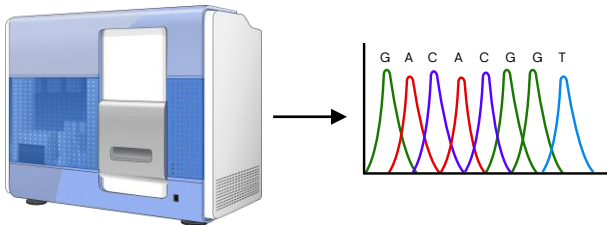
1. Extracción del ADN



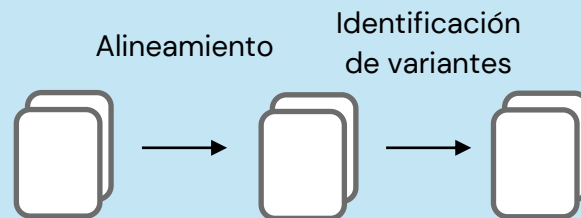
2. Preparación del ADN



3. Secuenciación



4. Análisis



COL7A1 - Nucleotide - NCBI

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

NIH

National Library of Medicine

National Center for Biotechnology Information

Log in

Nucleotide

Nucleotide

COL7A1

Search

Create alert

Advanced

Help

Species

Animals (1,368)

Bacteria (4)

Customize ...

Molecule types

genomic DNA/RNA (623)

mRNA (697)

Customize ...

Source databases

INSDC (GenBank) (534)

RefSeq (1,043)

Customize ...

Sequence Type

Nucleotide (1,572)

EST (4)

GSS (3)

Sequence length

Custom range...

Release date

Custom range...

Summary

20 per page

Sort by Default order

Send to:

Filters: [Manage Filters](#)

GENE

Was this helpful?

[COL7A1 – collagen type VII alpha 1 chain](#)

[Homo sapiens \(human\)](#)

Also known as: EBD1, EBDCT, EBR1, NDNC8

Gene ID: 1294

[RefSeq transcripts \(25\)](#)

[RefSeq proteins \(15\)](#)

[RefSeqGene \(1\)](#)

[PubMed \(202\)](#)

Orthologs

Genome Data Viewer

BLAST

RefSeq Sequences

Results by taxon

Top Organisms [Tree](#)

synthetic construct (138)

Homo sapiens (59)

Pan troglodytes (24)

Tyto alba (20)

Motacilla alba (19)

All other taxa (2265)

More...

Find related data

Database:

Select

Find items

Search details

COL7A1[All Fields]

25 selected items - Nucleotide

ncbi.nlm.nih.gov/nuccore/NM_000094.4,XM_017005688.2,XM_017005689.2,XM_017005690.2,XM_017005691.2,XM_017005692.2,XM_05...

NIH

National Library of Medicine

National Center for Biotechnology Information

Log in

Nucleotide

Nucleotide

Advanced

Search

Help

Species

Animals (1)

Customize ...

Molecule types

mRNA (1)

Customize ...

Source databases

RefSeq (1)

Customize ...

Sequence Type

Nucleotide (1)

Sequence length

Custom range...

Release date

Custom range...

Revision date

Custom range...

Clear all

Summary

20 per page

Sort by Default order

Send to:

Filters: [Manage Filters](#)

Items: 1 to 20 of 25

<< First < Prev Page 1 of 2 Next > Last >>

☐

[Homo sapiens collagen type VII alpha 1 chain \(COL7A1\), mRNA](#)

1. 9,231 bp linear mRNA

Accession: NM_000094.4 GI: 1812585486

[Protein](#) [PubMed](#) [Taxonomy](#)

[GenBank](#) [FASTA](#) [Graphics](#)

☐

[PREDICTED: Homo sapiens collagen type VII alpha 1 chain \(COL7A1\), transcript variant X1, mRNA](#)

2. 9,171 bp linear mRNA

Accession: XM_017005688.2 GI: 2217341740

[GenBank](#) [FASTA](#) [Graphics](#)

☐

[PREDICTED: Homo sapiens collagen type VII alpha 1 chain \(COL7A1\), transcript variant X5, mRNA](#)

3. 7,950 bp linear mRNA

Accession: XM_017005689.2 GI: 2217341744

Find related data

Database: Select

Find items

Recent activity

COL7A1 collagen type VII alpha 1 chain [Homo sapiens]

Homo sapiens collagen type VII alpha 1 chain (COL7A1), mRNA

COL7A1 (2525)

See more...



Nucleotide

Nucleotide

Advanced

Search

Help

GenBank

Homo sapiens collagen type VII alpha 1 chain (COL7A1), mRNA

NCBI Reference Sequence: NM_000094.4

[FASTA](#) [Graphics](#)

Go to:

LOCUS NM_000094 9231 bp mRNA linear PRI 15-FEB-2024
DEFINITION Homo sapiens collagen type VII alpha 1 chain (COL7A1), mRNA.
ACCESSION NM_000094 XM_011533337
VERSION NM_000094.4
KEYWORDS RefSeq; MANE Select.
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 9231)
AUTHORS Natale,M.I., Manzur,G.B., Lusso,S.B., Cella,E., Giovo,M.E.,

Send to:

- ☒ Complete Record
☐ Coding Sequences
☐ Gene Features

Choose Destination

- ☒ File ☐ Clipboard
☐ Collections ☐ Analysis Tool

Download 1 item.

Format

FASTA

Show GI

Create File

Articles about the COL7A1 gene

Trial of Beremagene Geperpavec (B-VEC) for
Dystrophic Epidermolysis Bi [N Engl J Med. 2022]


```
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', '')
```

```
    patient_file = askopenfile(mode='r')
    patient_seq = patient_file.readlines()[1:]
    patient_seq = ''.join(patient_seq).replace('\n', '')
```

```
    seq_
    for
```



marinamorolopez / alignment-WTMPython

```
        seq_alignment.append(pos_align)
```

```
    mutation_pos = [i for i, val in enumerate(seq_alignment) if not val]
    for i in range(len(mutation_pos)):
        original_bases = gene_seq[mutation_pos[i]]
        mutated_bases = patient_seq[mutation_pos[i]]
    mutation_pos_corrected = [x+1 for x in mutation_pos]
    print("Mutation positions: " + str(mutation_pos_corrected))
    print("Original bases: " + original_bases)
    print("Mutated bases: " + mutated_bases)
```

```
main()
```

```
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', '')

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

    seq_alignment = []
    for i in range(len(gene_seq)):
        pos_align = (gene_seq[i]==patient_seq[i])
        seq_alignment.append(pos_align)

    mutation_pos = [i for i, val in enumerate(seq_alignment) if not val]
    for i in range(len(mutation_pos)):
        original_bases = gene_seq[mutation_pos[i]]
        mutated_bases = patient_seq[mutation_pos[i]]
    mutation_pos_corrected = [x+1 for x in mutation_pos]
    print("Mutation positions: " + str(mutation_pos_corrected))
    print("Original bases: " + original_bases)
    print("Mutated bases: " + mutated_bases)

main()
```


Epidermólisis bullosa distrófica recesiva causada por T>C en posición 6591 del gen COL7A1 (entre otras)

TCTCTGGAGAACAGGGACCCCCTGGACTCAAGGGTGCTAAGGGGGAGCCGGGCAGCAATGGTGACCAAGG
TCCCAAAGGAGACAGGGGTGTGCCAGGCATCAAAGGAGACCGGGGAGAGCCTGGACCGAGGGGTTCAGGAC
GGCAACCCGGGTCTACCAGGAGAGCGTGGTATGGCTGGGCCTGAAGGGAAGCCGGGTCTGCAGGGTCCAA
GAGGCCCCCCCTGGCCCCAGTGGGTGGTCATGGAGACCCTGGACCACCTGGTGCCCCGGGTCTTGCTGGCCC
TGCAGGACCCCAAGGACCTTCTGGCCTGAAGGGGGAGCCTGGAGAGACAGGACCTCCAGGACGGGGCCTG
ACTGGACCTACTGGAGCTGTGGGACTTCCTGGACCCCCCGGCCCTTCAGGCCTTGTGGGTCCACAGGGGT
CTCCAGGTTTGCCTGGACAAGTGGGGGAGACAGGGAAGCCGGGAGCCCCAGGTTCGAGATGGTGCCAGTGG

TCTCTGGAGAACAGGGACCCCCTGGACTCAAGGGTGCTAAGGGGGAGCCGGGCAGCAATGGTGACCAAGG
TCCCAAAGGAGACAGGGGTGTGCCAGGCATCAAAGGAGACCGGGGAGAGCCTGGACCGAGGGGTTCAGGAC
GGCAACCCGGGTCTACCAGGAGAGCGTGGTATGGCTGGGCCTGAAGGGAAGCCGGGTCTGCAGGGTCCAA
GAGGCCCCCCCTCGGCCCAAGTGGGTGGTCATGGAGACCCTGGACCACCTGGTGCCCCGGGTCTTGCTGGCCC
TGCAGGACCCCAAGGACCTTCTGGCCTGAAGGGGGAGCCTGGAGAGACAGGACCTCCAGGACGGGGCCTG
ACTGGACCTACTGGAGCTGTGGGACTTCCTGGACCCCCCGGCCCTTCAGGCCTTGTGGGTCCACAGGGGT
CTCCAGGTTTGCCTGGACAAGTGGGGGAGACAGGGAAGCCGGGAGCCCCAGGTTCGAGATGGTGCCAGTGG

Epidermólisis bullosa distrófica recesiva causada por T>C en posición 6591 del gen COL7A1 (entre otras)

```
TCTCTGGAGAACAGGGACCCCCTGGACTCAAGGGTGCTAAGGGGGAGCCGGGCAGCAATGGTGACCAAGG
TCCCAAAGGAGACAGGGGTGTGCCAGGCATCAAAGGAGACCGGGGAGAGCCTGGACCGAGGGGTTCAGGAC
GGCAACCCGGGTCTACCAGGAGAGCGTGGTATGGCTGGGCCTGAAGGGAAGCCGGGTCTGCAGGGTCCAA
GAGGCCCCCCCTGGCCCCAGTGGGTGGTCATGGAGACCCTGGACCACCTGGTGCCCCGGGTCTTGCTGGCCC
TGCAGGACCCCAAGGACCTTCTGGCCTGAAGGGGGAGCCTGGAGAGACAGGACCTCCAGGACGGGGCCTG
ACTGGACCTACTGGAGCTGTGGGGCCTTGTGGGTCCACAGGGGTGGCCTTGTGGGTCCACAGGGGT
CTCCAGGTTTGCCTGGACAAGTCAGGTTCGAGATGGTGCCAGTGGCAGGTTCGAGATGGTGCCAGTGG

TCTCTGGAGAACAGGGACCCCCCGGGCAGCAATGGTGACCAAGG
TCCCAAAGGAGACAGGGGTGTGCCAGGCATCAAAGGAGACCGGGGAGAGCCTGGACCGAGGGGTTCAGGAC
GGCAACCCGGGTCTACCAGGAGAGCGTGGTATGGCTGGGCCTGAAGGGAAGCCGGGTCTGCAGGGTCCAA
GAGGCCCCCCCTCGGCCAGTGGGTGGTCATGGAGACCCTGGACCACCTGGTGCCCCGGGTCTTGCTGGCCC
TGCAGGACCCCAAGGACCTTCTGGCCTGAAGGGGGAGCCTGGAGAGACAGGACCTCCAGGACGGGGCCTG
ACTGGACCTACTGGAGCTGTGGGACTTCCTGGACCCCCCGGCCCTTCAGGCCTTGTGGGTCCACAGGGGT
CTCCAGGTTTGCCTGGACAAGTGGGGGAGACAGGGAAGCCGGGAGCCCCAGGTTCGAGATGGTGCCAGTGG
```


Mutation positions: [6591]
Original bases: T
Mutated bases: C



4

Más aplicaciones clínicas con genética y Python

Más aplicaciones




Descripción de
nuevas
mutaciones



Diseño de
tratamientos
genéticos



Estudio de
eficacia de
medicamentos



Predicción de
efectos de
mutaciones



Análisis de big
data clínico

¡Gracias! :)

¿Preguntas?



marinaml5598@gmail.com



marinamorolopez



Marina Moro López



marinamorolopez / alignment-WTMPython

