

Bioinformática - Introducción

Universidad de Ingeniería y Tecnología

PhD(c). Vicente Machaca Arceda

2022-I



Contenido

Introducción

 Presentación

 Objetivos

 Motivación

 Propósito de la Bioinformática

Bioinformática

 Definición

Aplicaciones

 Alineamiento y árboles filogenéticos

 Predicción de estructuras de proteínas

 Descubrimiento de nuevos medicamentos

 Inmunoterapia del Cáncer

Contenido

Introducción

Presentación

Objetivos

Motivación

Propósito de la Bioinformática

Bioinformática

Definición

Aplicaciones

Alineamiento y árboles filogenéticos

Predicción de estructuras de proteínas

Descubrimiento de nuevos medicamentos

Inmunoterapia del Cáncer

Presentación

- PhD(c). Vicente Enrique Machaca Arceda.

Presentación

- PhD(c). Vicente Enrique Machaca Arceda.
- Profesor UTEC (TP).

Presentación

- PhD(c). Vicente Enrique Machaca Arceda.
- Profesor UTEC (TP).
- Investigador en Bioinformática y Aprendizaje de Maquina.

Presentación

- PhD(c). Vicente Enrique Machaca Arceda.
- Profesor UTEC (TP).
- Investigador en Bioinformática y Aprendizaje de Maquina.
- Index-h 5.

Presentación

Publicaciones

Year	Country	Title
2020	USA	Small Ship Detection on Optical Satellite Imagery with YOLO and YOLT
2018	Brasil	Fast Car Crash Detection in Video
2016	Chile	Fast Face Detection in Violent Video Scenes
2016	Costa Rica	Real Time Violence Detection in Video with ViF and Horn-Schunck
2016	Costa Rica	Optimization model for face detection in video sequences
2015	Chile	Real Time Violence Detection in Video

Presentation

Publications

Year	Country	Title
2022	USA	ArgosMol: A Web Tool for Protein Structure Prediction and Visualization
2021	Chapter	COVID-19 Pandemic: Analysis and Statistics of Confirmed Cases
2020	Canada	An Analysis of k-Mer Frequency Features with Machine Learning Models for Viral Subtyping of Polyomavirus and HIV-1 Genomes
2020	Canada	An analysis of k-mer frequency features with SVM and CNN for viral subtyping classification
2020	Canada	Forecasting time series with Multiplicative Trend Exponential Smoothing and LSTM: COVID-19 case study

Contenido

Introducción

Presentación

Objetivos

Motivación

Propósito de la Bioinformática

Bioinformática

Definición

Aplicaciones

Alineamiento y árboles filogenéticos

Predicción de estructuras de proteínas

Descubrimiento de nuevos medicamentos

Inmunoterapia del Cáncer

Objetivos

- Conocer que es la Bioinformática y como se relaciona con las Ciencias de la Computación.
- Conocer las principales aplicaciones de la Bioinformática.

Contenido

Introducción

Presentación

Objetivos

Motivación

Propósito de la Bioinformática

Bioinformática

Definición

Aplicaciones

Alineamiento y árboles filogenéticos

Predicción de estructuras de proteínas

Descubrimiento de nuevos medicamentos

Inmunoterapia del Cáncer

Muertes por COVID-19

213 mil muertes por COVID en Perú [1].

6.29 millones de muertes por COVID en el mundo [1].

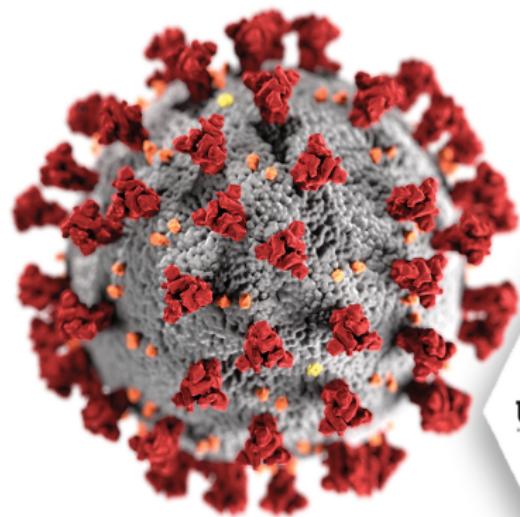


Figure: Visualización 3D del virus COVID-19.

Muertes por Cáncer

3.35 millones muertes por Cáncer en el mundo (2022) [2].

9.89 millones de muertes por Cáncer en el mundo [2].



Figure: Muertes por Cáncer en Perú en el 2020.

Reflexión

La Bioinformática representa un factor clave para el control de estas enfermedades.

Contenido

Introducción

Presentación

Objetivos

Motivación

Propósito de la Bioinformática

Bioinformática

Definición

Aplicaciones

Alineamiento y árboles filogenéticos

Predicción de estructuras de proteínas

Descubrimiento de nuevos medicamentos

Inmunoterapia del Cáncer

Propósito de la Bioinformática

¿Porque nos enfermamos de Cáncer?

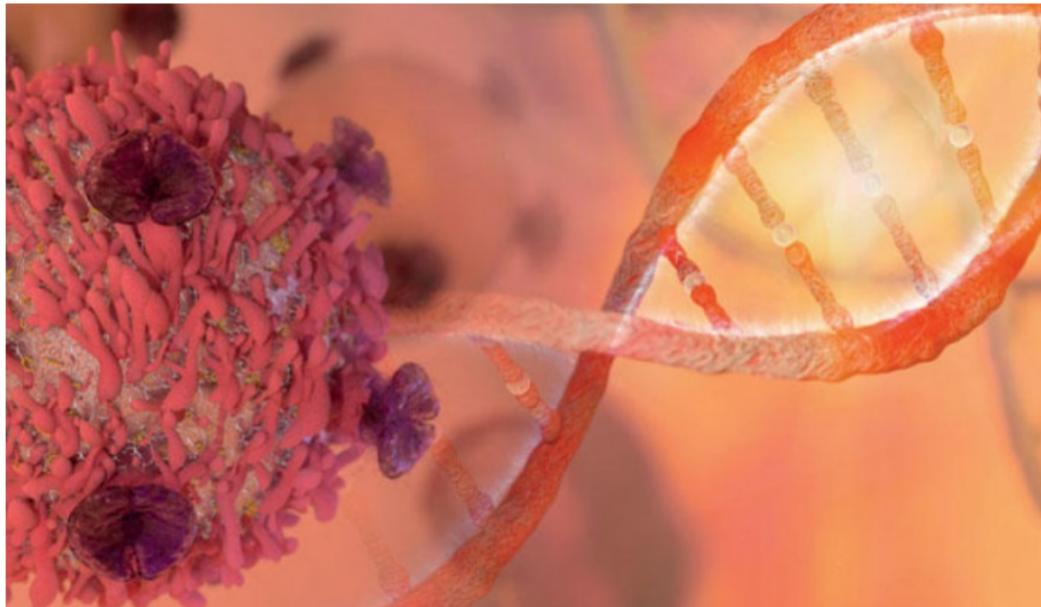


Figure: ¿Porque una persona se enferma de Cáncer?

Propósito de la Bioinformática

Descubrimiento de nuevos medicamentos

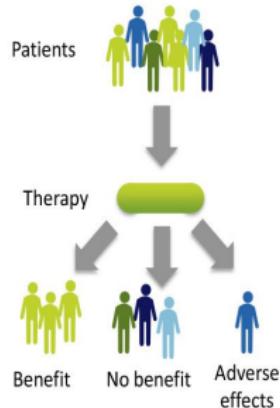


Figure: Descubrimiento de nuevos medicamentos.

Propósito de la Bioinformática

Medicina personalizada

Without Personalized Medicine:
Some Benefit, Some Do Not



With Personalized Medicine:
Each Patient Receives the Right Medicine For Them

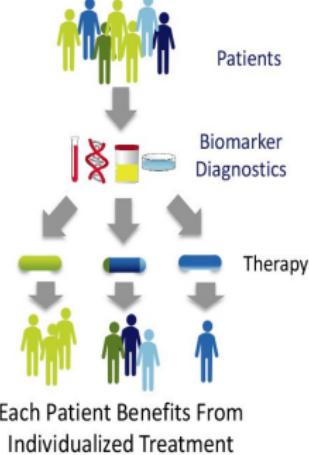


Figure: Medicina personalizada.

Propósito de la Bioinformática

¿Existe un gen de la bondad?



Figure: ¿Existe un gen de la bondad?

Contenido

Introducción

Presentación

Objetivos

Motivación

Propósito de la Bioinformática

Bioinformática

Definición

Aplicaciones

Alineamiento y árboles filogenéticos

Predicción de estructuras de proteínas

Descubrimiento de nuevos medicamentos

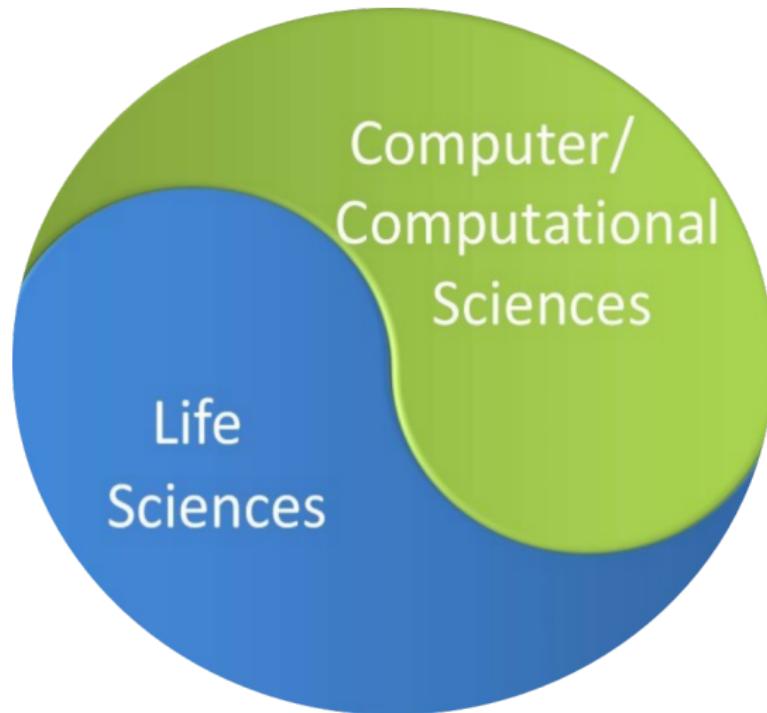
Inmunoterapia del Cáncer

Bioinformática

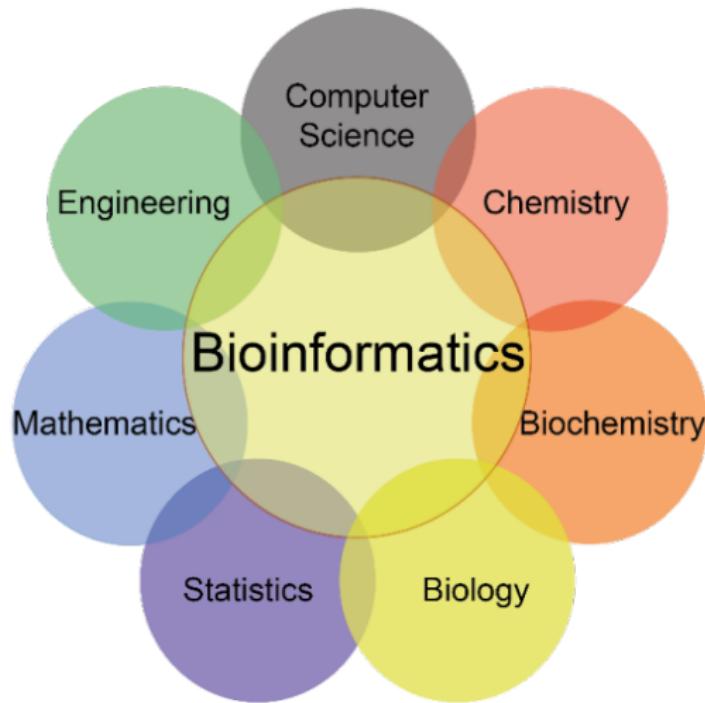
Definición

La Bioinformática involucra el uso de la tecnología de las computadoras para almacenar, recuperar, manipular y distribuir información relacionada a Biología Molecular como: **DNA, RNA, y proteínas** [3].

Bioinformática

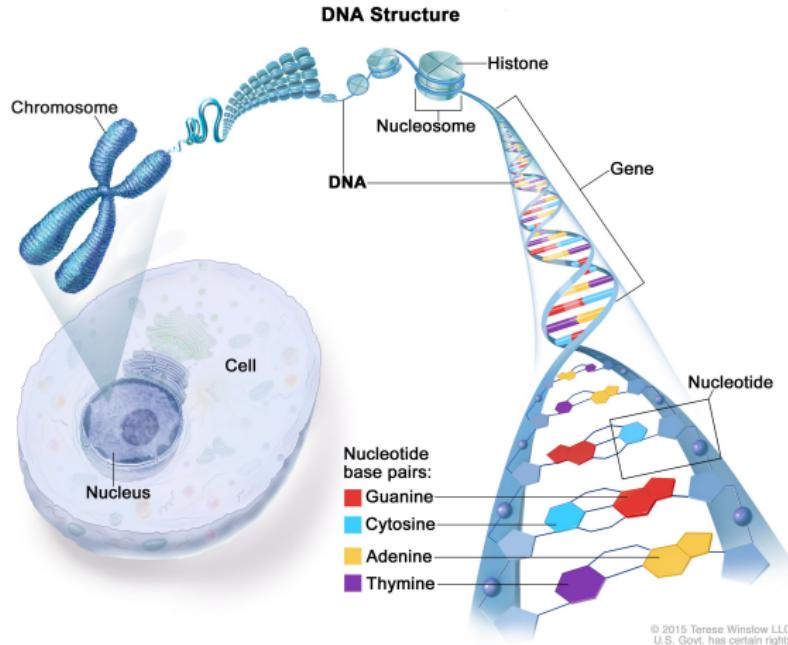


Bioinformática



Bioinformática

¿Donde está el DNA?



Bioinformática

Ejemplo de una secuencia de DNA

```
>J01859.1 Escherichia coli 16S ribosomal RNA, complete sequence
AAATTGAAGAGTTGATCATGGCTCAGATTGAACGCTGGCGGCAGGCCTAACACATGCAAGTCGAACGGT
AACAGGAAGAAGCTTGCTCTTGCTGACGAGTGGCGGACGGGTGAGTAATGTCCTGGGAAACTGCCTGATG
GAGGGGGATAACTACTGGAAACCGTGCTAATACCGCATAACGTCGCAAGACCAAAGAGGGGGACCTTCG
GGCCTTGTGCCATCGGATGTGCCAGATGGGATTAGCTAGTAGGTGGGGTAACGGCTCACCTAGGGCAGC
ATCCCTAGCTGGTCTGAGAGGTGACCAGGCACACTGGAACTGAGACACGGTCCAGACTCCTACGGGAGG
CAGCAGTGGGGAAATTGCAATGGCGCAAGCCTGATGCAGCCATGCCGCTGTATGAAGAAGGCCCTT
CGGGTTGTAAGTACTTCACTGGGGAGGAAGGGAGTAAAGTTAACCTTGCTATTGACGTTACCCG
CAGAAGAAGCACCGGCTAACCGTGCAGCAGCCGGTAATACGGAGGGTGCAAGCGTTAACCGGAA
TACTGGCGTAAAGCGCACCGAGCGTTTGTAGTCAGATGTGAAATCCCCTGGCTAACCTGGGAAAC
TGCACTGATACTGGCAAGCTTGAGTCTGTTAGAGGGGGTAGAATTCAAGGTGTAGCGGTAAATGCGT
AGAGATCTGGAGGAATACCGGTGGCGAAGGCGCCCCCTGGACGAAAGACTGACGCTCAGGTGCGAAAGCG
TGGGGAGCAACAGGATTAGATAACCTGGTAGTCACGCCGTAACCGATGTCGACTTGGAGGGTGTGCC
TTGAGGCGTGGCTTCCGGAGCTAACCGGTTAACGCGTTAGTCGACCGCCCTGGGGAGTACGGCCGCAAGGTTAAACT
CAAATGAATTGACGGGGGCCGCACAAGCGGTGGAGCATGTGGTTAACCGTACACCGCAAGAACCT
TACCTGGTCTTGACATCCACCGGAAGTTTCAGAGATGAGAATGTCCTCGGGAACCGTGAGACAGGTGC
TGCATGGCTGTCGTCACTCGTGTGAAATGTTGGGTTAAGTCCCGCAACGAGCGCAACCCCTATCCT
TTGTTGCCAGCGGTCCGGCCGGAACTCAAAGGAGACTGCCAGTGATAAAACTGGAGGAAGGTGGGGATGA
CGTCAAGTCATCATGGCCCTTACGACCAGGGCTACACACGTGCTACAATGGCGCATACAAGAGAAGCGA
CCTCGCGAGAGCAAGCGGACCTCATAAAGTGCCTGCTAGTCCGGATTGGAGTCTGCAACTCGACTCCATG
AAGTCGGAATCGCTAGTAATCGTGGATCAGAATGCCACGGTGAATACTGGTCCGGGCTTGTACACACCG
CCCGTCACACCATGGGAGTGGGTTGCAAAAGAAGTAGGTAGCTAACCTTCGGGAGGGCGCTTACCACTT
TGTGATTGACTGGGGTGAAGTCGTAACAAGGTAAACCGTAGGGGAACCTGCGGTTGGATCACCTCCTT
```

Figure: 16S ribosomal DNA of Escherichia coli with FASTA Format.

Bioinformática

Algunos números

Un genoma completo tiene **~3.2 billones de bp.**
~6.4 billones de bases [5].

Bioinformática

Algunos números

Un genoma completo tiene **~3.2 billones de bp.**
~6.4 billones de bases [5].

El genoma del virus **HIV-1** tiene aproximadamente **~20k bp.**
El genoma del virus **COVID-19** tiene aproximadamente **~32k bp [6].**

Bioinformática

Algunos números

Un genoma completo tiene **~3.2 billones de bp.**
~6.4 billones de bases [5].

El genoma del virus **HIV-1** tiene aproximadamente **~20k bp.**
El genoma del virus **COVID-19** tiene aproximadamente **~32k bp [6].**

Existe aproximadamente **19000 a 25000 genes [5].**

Bioinformática

Transcripción y Traducción

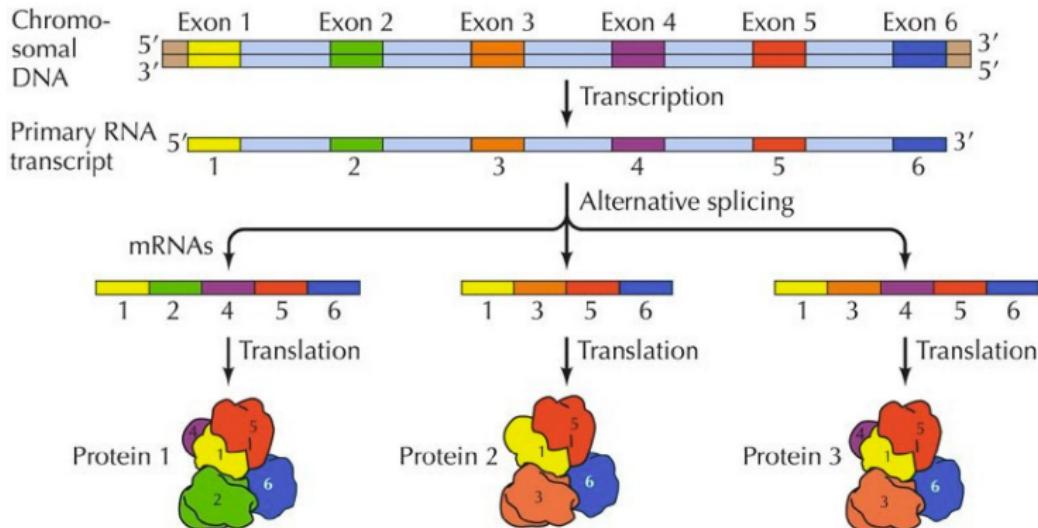


Figure: Alternative splicing [7].

Contenido

Introducción

Presentación

Objetivos

Motivación

Propósito de la Bioinformática

Bioinformática

Definición

Aplicaciones

Alineamiento y árboles filogenéticos

Predicción de estructuras de proteínas

Descubrimiento de nuevos medicamentos

Inmunoterapia del Cáncer

Aplicaciones de la Bioinformática

Orígenes del COVID-19

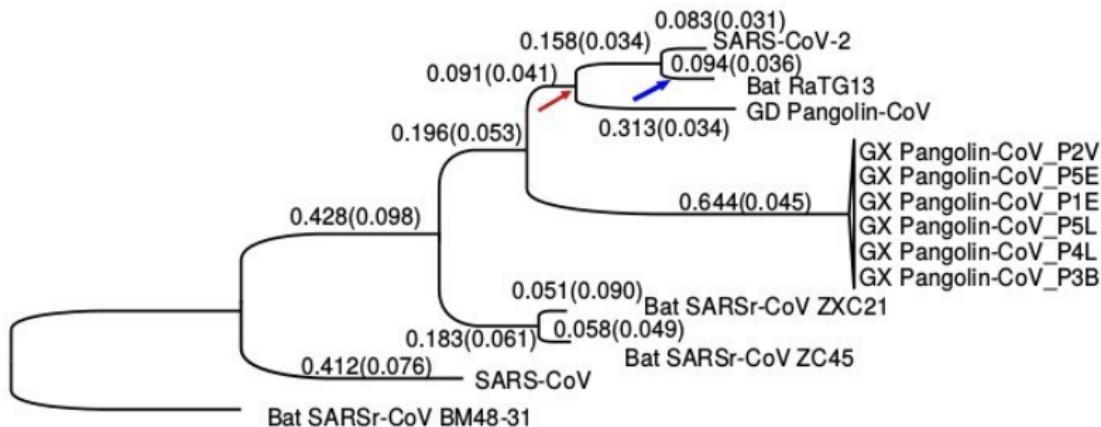


Figure: Árbol filogenético del COVID-19 y otros coronavirus [8].

Aplicaciones de la Bioinformática

Métodos alternativos de alineamiento

An Analysis of k-Mer Frequency Features with Machine Learning Models for Viral Subtyping of Polyomavirus and HIV-1 Genomes [9].

An analysis of k-mer frequency features with SVM and CNN for viral subtyping classification [10].

Contenido

Introducción

Presentación

Objetivos

Motivación

Propósito de la Bioinformática

Bioinformática

Definición

Aplicaciones

Alineamiento y árboles filogenéticos

Predicción de estructuras de proteínas

Descubrimiento de nuevos medicamentos

Inmunoterapia del Cáncer

Aplicaciones de la Bioinformática

Predicción de estructuras de proteínas

GSIGAASMEFCFDVFKELKVHHANENIFY
CPIAIMSALAMVYLGAKDSTRTQINKVVR
FD
KLPGFGDSIEAQCGTSVN VHSSLRDILNQI
TKPNDVYSFSLASRLYAEERYPILPEYLQC
VKELYRGGLEPINFTAADQARELINSWV
ESQTNGIRNVLPQSSVDSQTAMVLVNAIV
F
KGLWEKAFKDEDTQAMPFRVTEQESKPV
QMMYQIGLFRVASMASEKMKILELPFASG
TMS
MLVLLPDEVSGLEQLESIINFEKLTEWTSS
NVMEERKIKVYLPRMKMEEKYNLTSVLM
AM
GITDVFSANLSGISSAESLKISQAVHAAH
AEINEAGREVVGSAEAGVDAASVSEEFRA

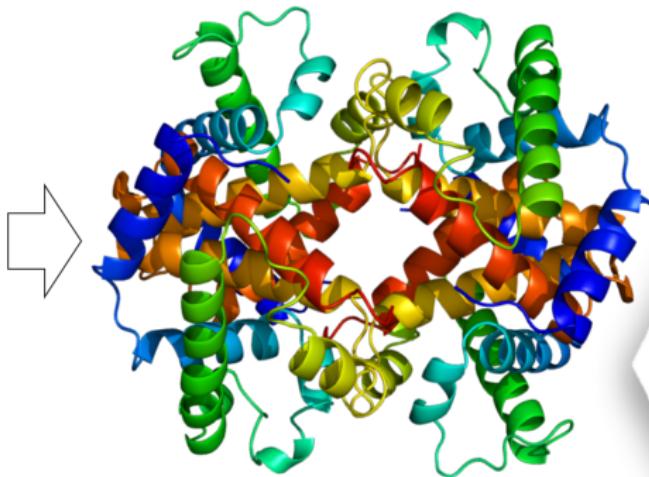


Figure: Simulación 3D de la estructura terciaria de una proteína.

Protein structure prediction

Método propuesto por AlphaFold

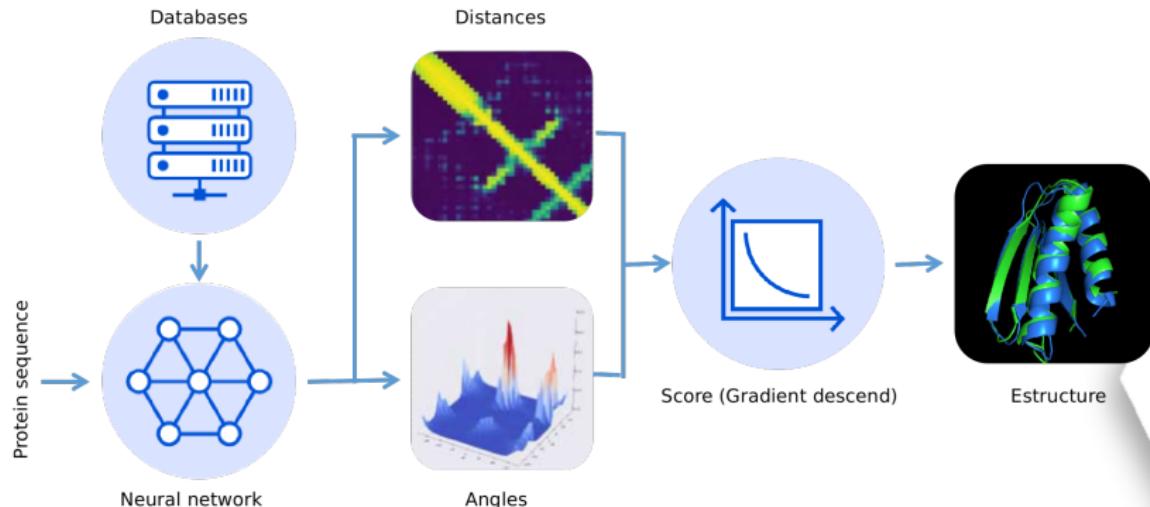


Figure: Predicción de estructuras de proteínas con AlphaFold [11].

Protein structure prediction

ArgosMol

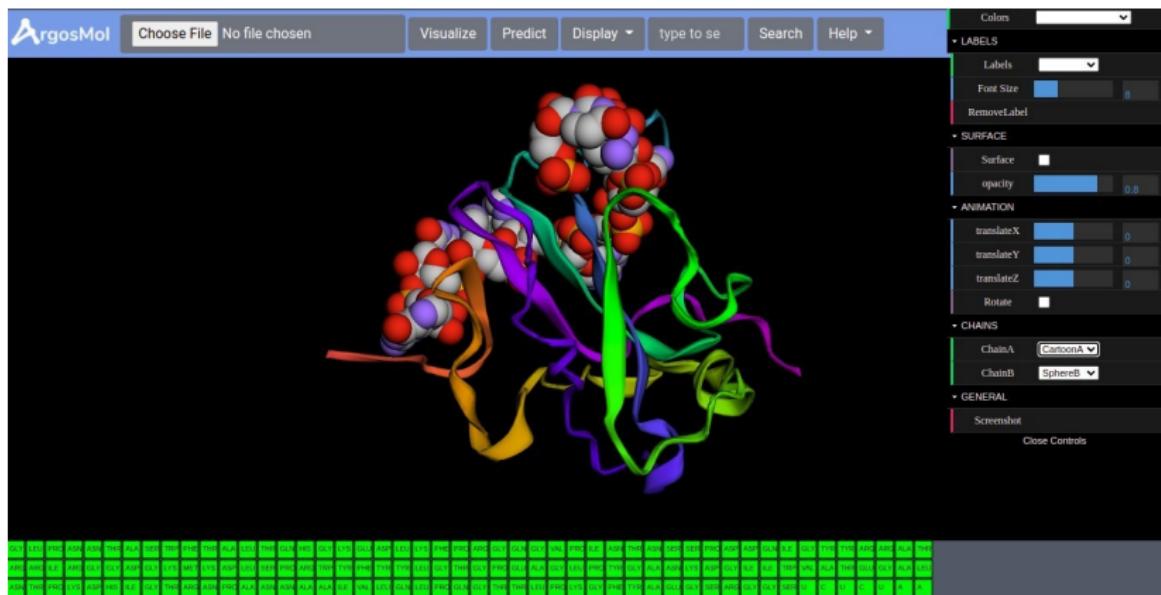


Figure: ArgosMol: <http://143.244.188.93:8080> [12].

Contenido

Introducción

Presentación

Objetivos

Motivación

Propósito de la Bioinformática

Bioinformática

Definición

Aplicaciones

Alineamiento y árboles filogenéticos

Predicción de estructuras de proteínas

Descubrimiento de nuevos medicamentos

Inmunoterapia del Cáncer

Aplicaciones de la Bioinformática

De un millón a uno

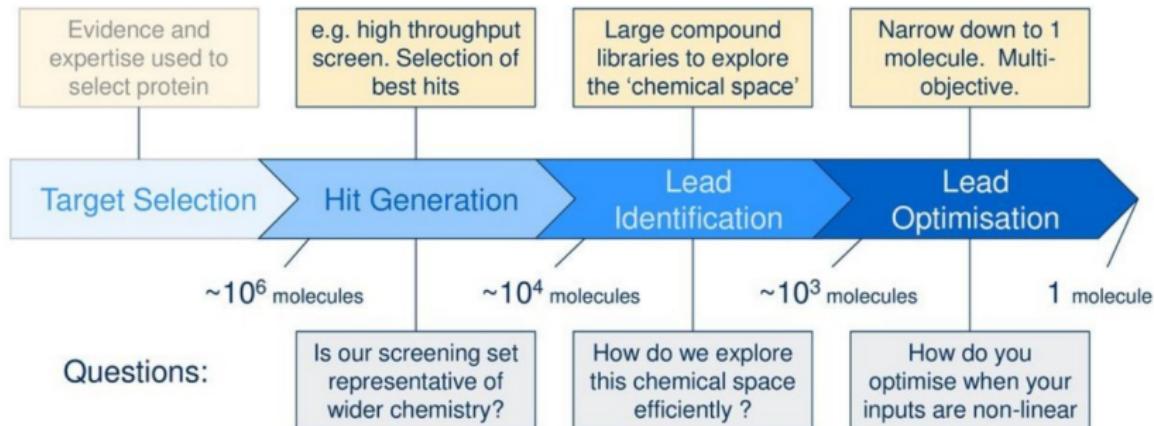


Figure: Proceso del descubrimiento de medicamentos [13].

Aplicaciones de la Bioinformática

Descubrimiento de nuevos medicamentos

Target



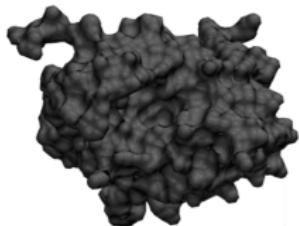
Ligand



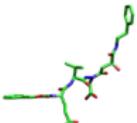
Complex



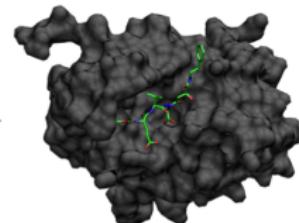
docking



+



docking



Aplicaciones de la Bioinformática

Descubrimiento de nuevos medicamentos

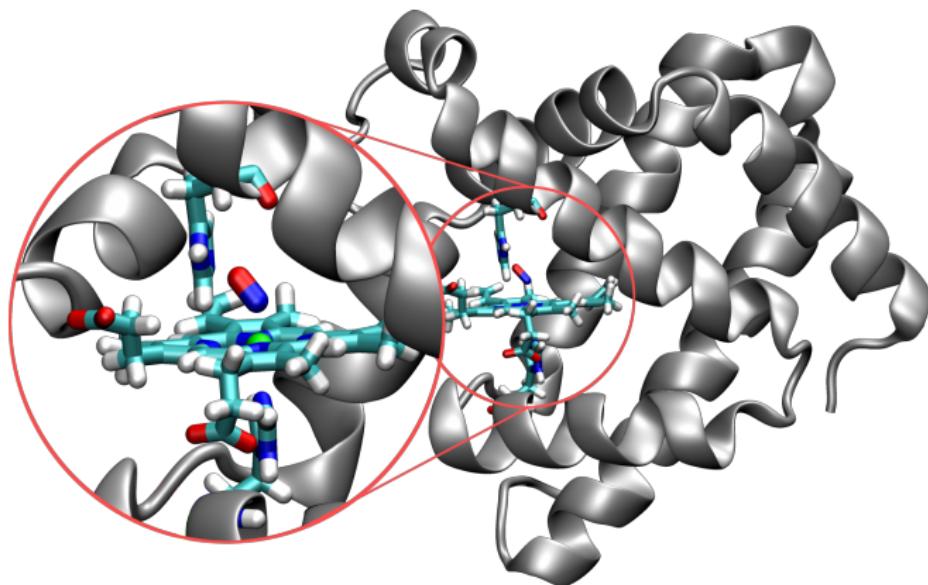


Figure: Enlace entre una molécula y una proteína.

Contenido

Introducción

Presentación

Objetivos

Motivación

Propósito de la Bioinformática

Bioinformática

Definición

Aplicaciones

Alineamiento y árboles filogenéticos

Predicción de estructuras de proteínas

Descubrimiento de nuevos medicamentos

Inmunoterapia del Cáncer

Aplicaciones de la Bioinformática

Inmunoterapia del Cáncer

Inmunoterapia del cáncer es un tipo de tratamiento contra el cáncer que estimula las defensas naturales del cuerpo para combatir el cáncer [14].

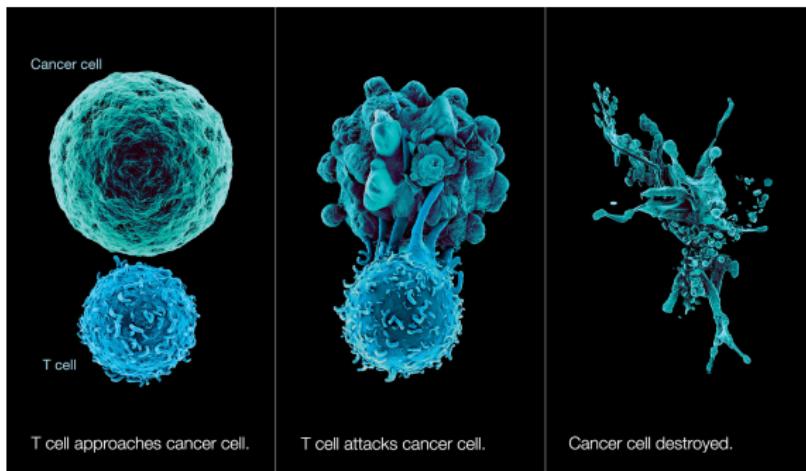


Figure: Ejemplo de como una célula T, destruye células del cancer.

Aplicaciones de la Bioinformática

Neo antígenos

Antígeno es cualquier sustancia que provoca que el sistema inmunitario produzca anticuerpos contra sí mismo [15].

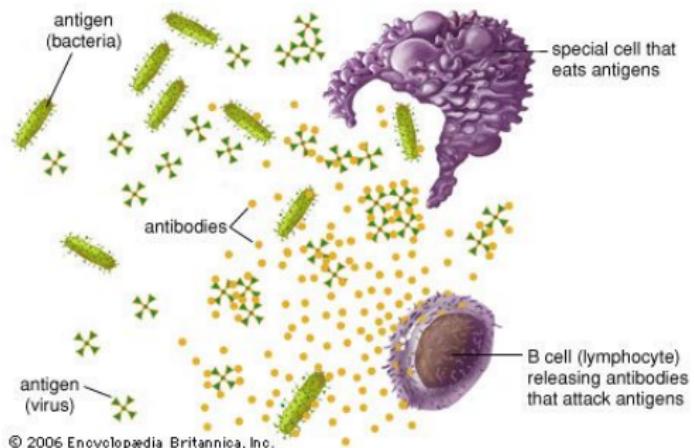


Figure: Ejemplos de antígenos y como son destruidos por el sistema inmune.
Fuente: [enlace](#).

Aplicaciones de la Bioinformática

Proceso

Neo antígeno es una proteína que se produce cuando aparecen ciertas mutaciones en el ADN de un tumor [16].

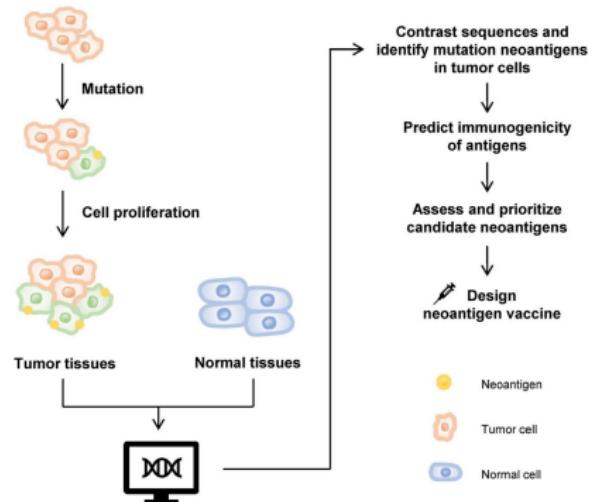


Figure: Proceso para la generación de vacunas personalizadas contra el cáncer. Fuente: [17].

Aplicaciones de la Bioinformática

Continuara...

References I

-  Worldometer, "Covid statistics," <https://www.worldometers.info/coronavirus/>, 2022, accessed: 2022-05-29.
-  ——, "Cancer statistics," <https://www.worldometers.info/cancer/>, 2022, accessed: 2022-05-29.
-  N. M. Luscombe, D. Greenbaum, and M. Gerstein, "What is bioinformatics? a proposed definition and overview of the field," *Methods of information in medicine*, vol. 40, no. 04, pp. 346–358, 2001.
-  NCI. (2022) National cancer institute dictionary. [Online]. Available: <https://www.cancer.gov/publications/dictionaries/genetics-dictionary>
-  J. M. Archibald, *Genomics: A Very Short Introduction*. Oxford University Press, 2018, vol. 559.

References II

-  G. S. Randhawa, M. P. Soltysiak, H. El Roz, C. P. de Souza, K. A. Hill, and L. Kari, "Machine learning using intrinsic genomic signatures for rapid classification of novel pathogens: Covid-19 case study," *bioRxiv*, 2020.
-  G. BIO, "Gen. bio,"
<https://sites.google.com/site/bio1040genbio2/home>, 2020,
accessed: 2020-03-20.
-  X. Tang, C. Wu, X. Li, Y. Song, X. Yao, X. Wu, Y. Duan, H. Zhang, Y. Wang, Z. Qian *et al.*, "On the origin and continuing evolution of sars-cov-2," *National Science Review*, 2020.
-  V. M. Arceda, "An analysis of k-mer frequency features with machine learning models for viral subtyping of polyomavirus and hiv-1 genomes," in *Proceedings of the Future Technologies Conference*. Springer, 2020, pp. 279–290.

References III

-  V. E. Machaca Arceda, "An analysis of k-mer frequency features with svm and cnn for viral subtyping classification," *Journal of Computer Science & Technology*, vol. 20, 2020.
-  A. Senior, J. Jumper, D. Hassabis, and P. Kohli, "AlphaFold: Using ai for scientific discovery," 2020.
-  E. S. Condori, J. S. Lupa, S. B. Cornejo, and V. M. Arceda, "Argosmol: A web tool for protein structure prediction and visualization," in *Future of Information and Communication Conference*. Springer, 2022, pp. 604–616.
-  C. Nantasesamat, "Computational drug discovery: Machine learning for making sense of big data in drug discovery," 2020.
-  Cancer.net. (2022) Qué es la inmunoterapia. [Online]. Available: <https://www.cancer.net/es/desplazarse-por-atencion-del-cáncer/como-se-trata-el-cáncer/inmunoterapia/qué-es-la-inmunoterapia>

References IV

-  Medline.plus. (2022) Antígeno. [Online]. Available: <https://medlineplus.gov/spanish/ency/article/002224.htm>
-  I. N. del Cancer. (2022) Neo antígeno. [Online]. Available: <https://www.cancer.gov/espanol/publicaciones/diccionarios/diccionario-cancer/def/neoantigeno>
-  M. Peng, Y. Mo, Y. Wang, P. Wu, Y. Zhang, F. Xiong, C. Guo, X. Wu, Y. Li, X. Li *et al.*, “Neoantigen vaccine: an emerging tumor immunotherapy,” *Molecular cancer*, vol. 18, no. 1, pp. 1–14, 2019.

¿Preguntas?

