

Bioinformatics against COVID-19

Bioinformatics Research Group

MSc. Vicente Machaca Arceda

August 18, 2020

Overview



Introduction

Presentation

The purpose of Bioinformatics

What is Bioinformatics?

Bioinformatics against COVID-19

COVID origin

Protein structure prediction

Drug discovery

Funding

Datasets and resources

How to learn?

How to learn Bioinformatics?

Table of Contents



Introduction

Presentation

The purpose of Bioinformatics

What is Bioinformatics?

Bioinformatics against COVID-19

COVID origin

Protein structure prediction

Drug discovery

Funding

Datasets and resources

How to learn?

How to learn Bioinformatics?

Presentation



- ▶ MSc. Vicente Enrique Machaca Arceda.

Presentation



- ▶ MSc. Vicente Enrique Machaca Arceda.
- ▶ Professor at UNSA university.

Presentation



- ▶ MSc. Vicente Enrique Machaca Arceda.
- ▶ Professor at UNSA university.
- ▶ Full time researcher at La Salle university.

Presentation



- ▶ MSc. Vicente Enrique Machaca Arceda.
- ▶ Professor at UNSA university.
- ▶ Full time researcher at La Salle university.
- ▶ Leader of Bioinformatics Research Group in Arequipa.

Presentation

Publications



| Year | Country | Title |
|------|------------|---|
| 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 |
|------|---------|--|
| 2020 | | DNA sequence similarity analysis using Chaos Game Representation |
| 2020 | | Machine Learning and Chaos Game Representation for rapid classification of novel pathogens COVID-19 case study |
| 2020 | Canada | An analysis of k-mer frequency features with machine learning models for viral subtyping of Polyomavirus and HIV-1 genomes |
| 2020 | Canada | Forecasting time series with Multiplicative Trend Exponential Smoothing and LSTM: COVID-19 case study |
| 2020 | USA | Small Ship Detection on Optical Satellite Imagery with YOLO and YOLT |

Table of Contents



Introduction

Presentation

The purpose of Bioinformatics

What is Bioinformatics?

Bioinformatics against COVID-19

COVID origin

Protein structure prediction

Drug discovery

Funding

Datasets and resources

How to learn?

How to learn Bioinformatics?

The purpose of Bioinformatics

Why a person has cancer?

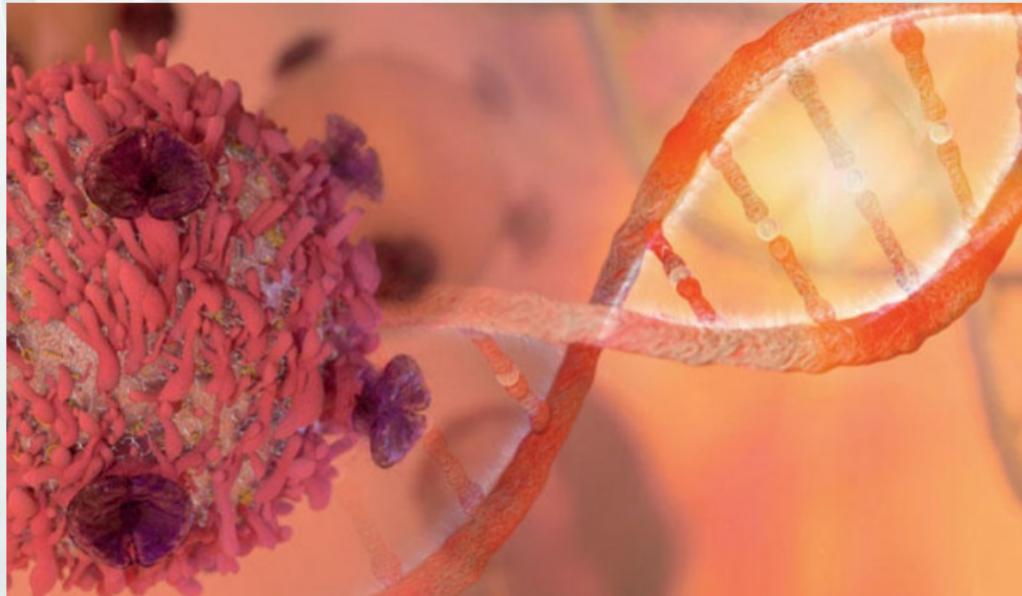


Figure: Why a person has cancer?

The purpose of Bioinformatics

Why some medicines no work in some persons?



Figure: Why some medicines no work in some persons?

The purpose of Bioinformatics

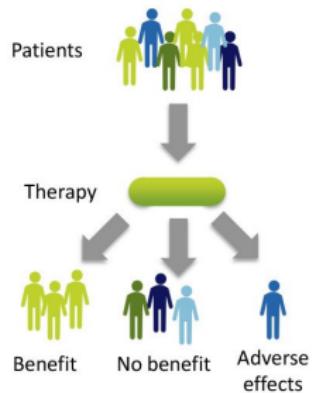
Treatment Development



9

Without Personalized Medicine:

Some Benefit, Some Do Not



With Personalized Medicine:

Each Patient Receives the Right Medicine For Them

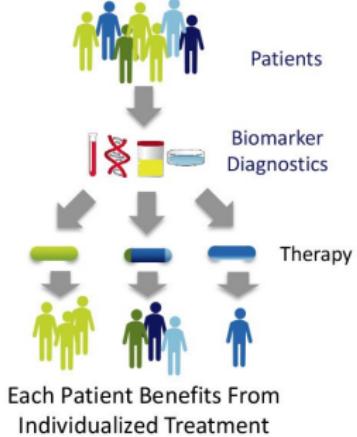


Figure: Personalized Medicine: New Approach to Treatment of Disease

The purpose of Bioinformatics

Protein structure prediction



10

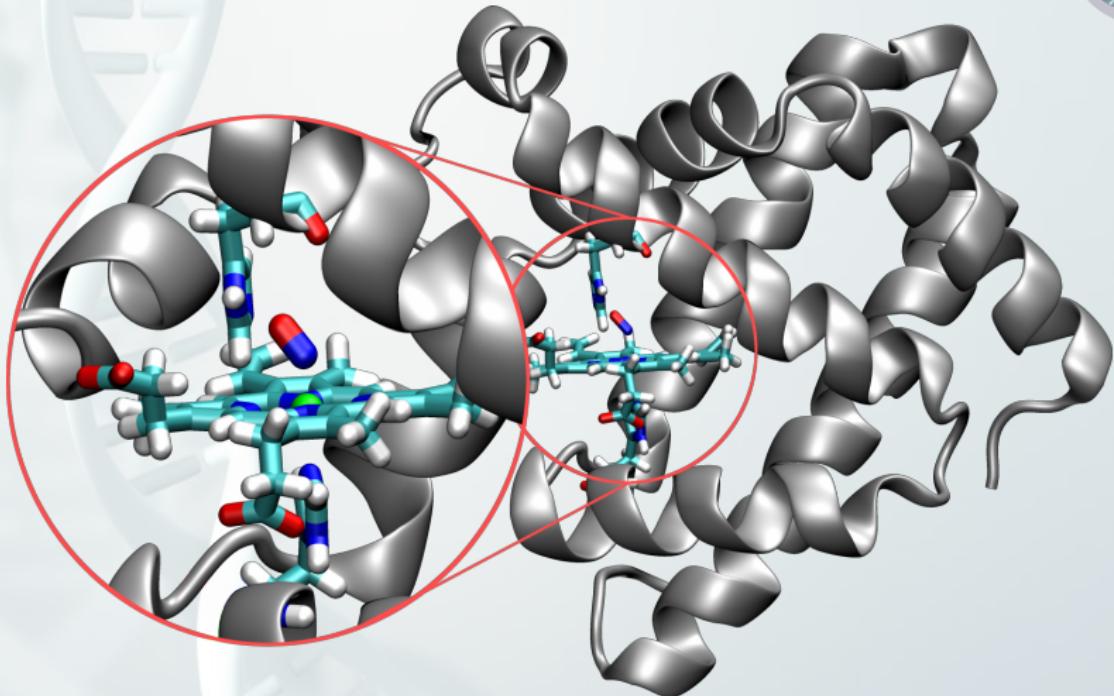


Figure: Computer simulation of protein-ligand.

Table of Contents



11

Introduction

Presentation

The purpose of Bioinformatics

What is Bioinformatics?

Bioinformatics against COVID-19

COVID origin

Protein structure prediction

Drug discovery

Funding

Datasets and resources

How to learn?

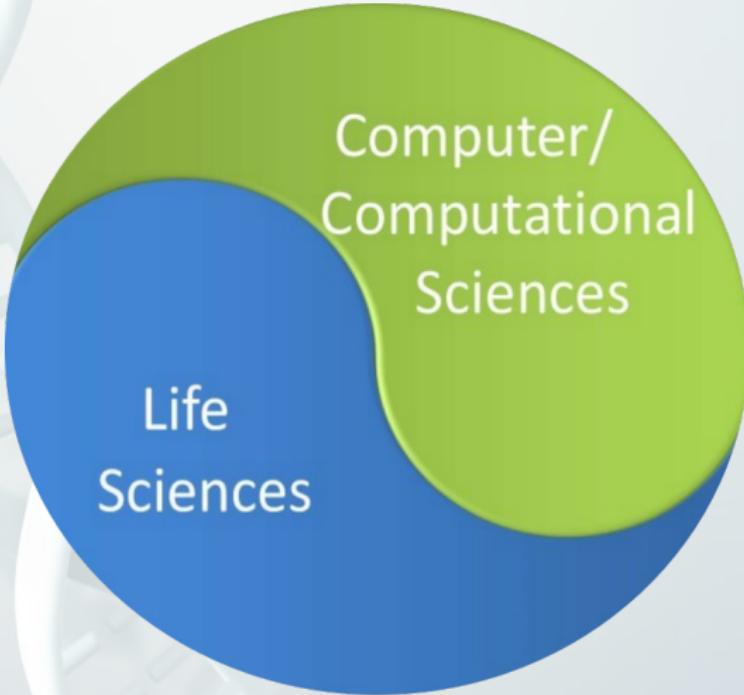
How to learn Bioinformatics?

Introduction

What is Bioinformatics?



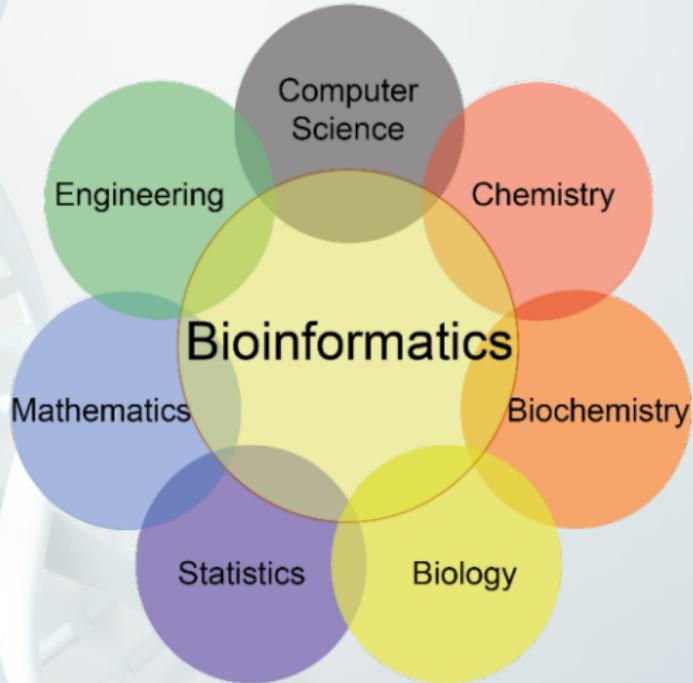
According to Luscombe et al.: **Bioinformatics** involves the technology that uses computers for storage, retrieval, manipulation, and distribution of information related to biological macromolecules such as DNA, RNA, and proteins [1].



Bioinformatics



14



Bioinformatics

Where is DNA located?



15

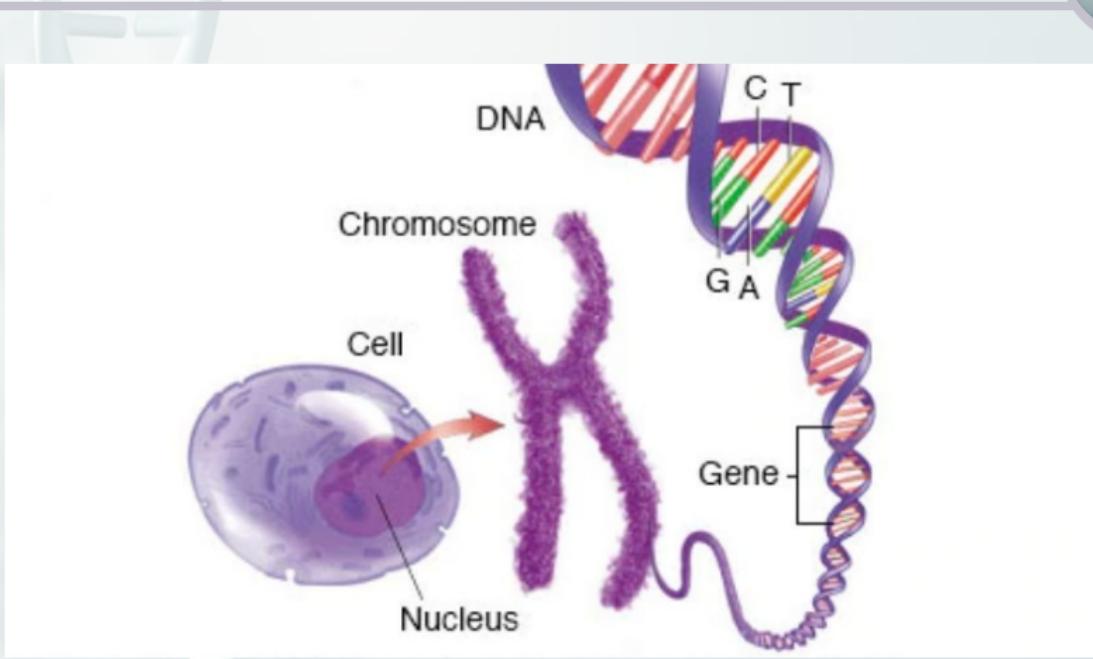


Figure: Where DNA is located [2].

Bioinformatics

DNA structure

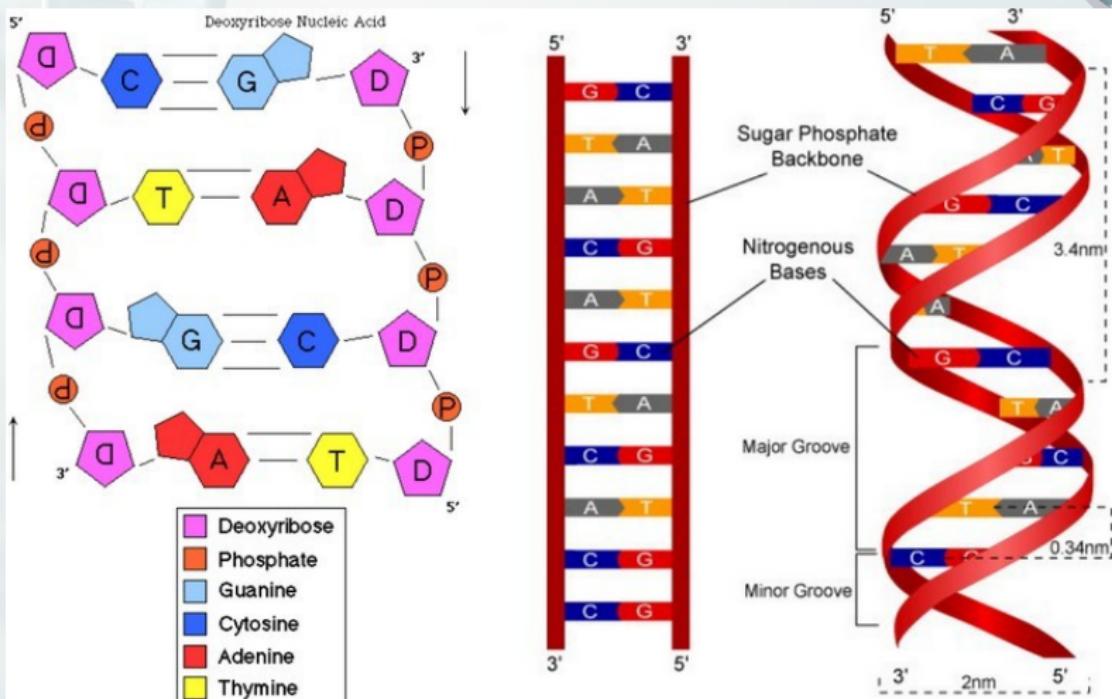


Figure: DNA structure [3].



The human genome is made of **~3.2 billions bp** of DNA.
~6.4 billions of nucleotides [4].



The human genome is made of **~3.2 billions bp** of DNA.
~6.4 billions of nucleotides [4].

The HIV-1 genome is made of **~20k bp** of DNA.
Meanwhile, the COVID-19 is made of **~32k bp** [5].



The human genome is made of **~3.2 billions bp** of DNA.
~6.4 billions of nucleotides [4].

The HIV-1 genome is made of **~20k bp** of DNA.
Meanwhile, the COVID-19 is made of **~32k bp** [5].

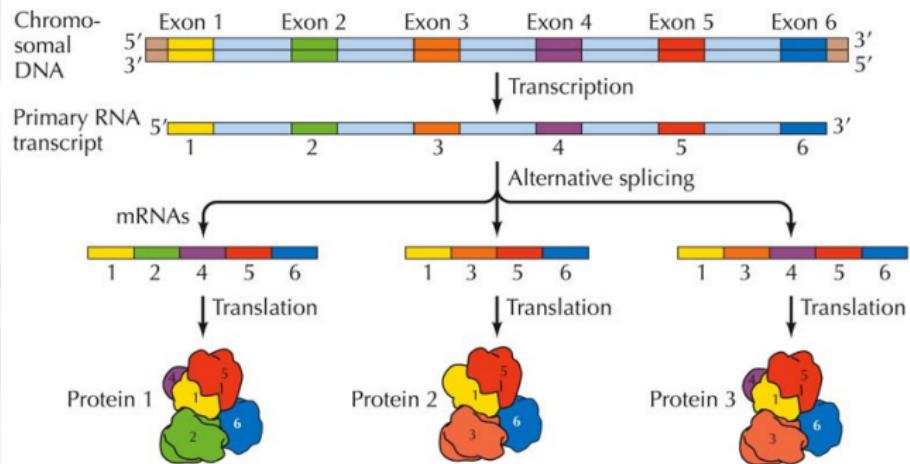
There are approximately **19000 to 25000** genes.
No one knows for sure [4].

Bioinformatics

Transcription and translation



18



THE CELL, Fourth Edition, Figure 5.5 © 2006 ASM Press and Sinauer Associates, Inc.

Figure: Alternative splicing [6].

Bioinformatics

Example of DNA sequence



19

```
>J01859.1 Escherichia coli 16S ribosomal RNA, complete sequence
AAATTGAAGAGTTGATCATGGCTCAGATTGAACGCTGGCGCAGGCCTAACACATGCAAGTCGAACGGT
AACAGGAAGAACGCTTGTCTTTGTCGAGTGCGGACGGGTGAGTAATGTCGGAAACTGCCTGATG
GAGGGGGATAACTACTGAAACGGTAGCTAACCGCATAACGTCGCAAGACCAAAGAGGGGGACCTTCG
GGCCTCTGCCATCGGATGTGCCAGATGGGATTAGCTAGTAGGTGGGGTAACGGCTCACCTAGGCACG
ATCCCTAGCTGGTCTGAGAGGATGACCAGCCACACTGGAACGTGAGACACGGTCCAGACTCTACGGGAGG
CAGCAGTGGGAATATTGCAACATGGCGCAAGCCTGATGCAGGCCATGCCGCGTGTATGAAGAACGGCTT
CGGGTTGTAAGTACTTCAGCGGGAGGAAGGGAGTAAAGTTAACACCTTGCTCATTGACGTTACCCG
CAGAAGAACGACCGGCTAACCTCGTGCAGCAGCCGCGTAAACGGAGGGTCAAGCGTTAACCGGAAT
TACTGGGCGTAAAGCGCACGCAGCGGTTGTTAAGTCAGATGTGAAATCCCCGGCTAACCTGGGAC
TGCATCTGATACTGGCAAGCTTGAGTCTCGTAGAGGGGGTAGAAATTCCAGGTGAGCGGTGAAATGCGT
AGAGATCTGGAGGAATACCGGTGGCGAAGCGGCCCTGGACGAAGACTGACGCTCAGGTGCGAACGCG
TGGGGAGCAAACAGGATTAGATAACCTGGTAGTCACGCCGTAACGATGTCGACTTGGAGGTTGTGCC
TTGAGGCGTGGCTTCCGGAGCTAACCGCTTAAGTCGACGCCCTGGAGTACGGCGCAAGGTTAAACT
CAAATGAATTGACGGGGGCCCGACAAGCGGTGGAGCATGTTAACCGATGCAACCGAACGCGAAGAACCT
TACCTGGTCTTGACATCACGGAAAGTTTCAGAGATGAGAAATGTCGCTTGGGAACCGTGAGACAGGTG
TGCATGGCTGTCGTCAGCTCGTGTGAAATGTTGGGTTAACGCGCAACGCCCTATCCT
TTGTTGCCAGCGGTCGCCGGGAACTCAAAGGAGACTGCCAGTGATAAACTGGAGGAAGGTGGGGATG
CGTCAAGTCATCATGGCCCTTACGACCAGGGCTACACACGTCTACAATGGCGCATACAAGAGAACGCG
CCTCGCGAGAGCAAGCGGACCTCATAAAGTGCCTGCTAGTCCGGATTGGAGTCTGCAACTCGACTCCATG
AAGTCGGAATCGCTAGTAATCGTGGATCAGAAATGCCACGGTGAATACGTTCCGGGCTTGTACACACCG
CCCGTCACACCATGGGAGTGGGTTGCAAAAGAAGTAGGTAGCTAACCTTGGAGGGCGCTTACCACTT
TGTGATTGACTGGGGTGAAGTCGTAACAAGGTAAACCGTAGGGAACCTGCGGTTGGATCACCTCCTT
```

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

Table of Contents



Introduction

Presentation

The purpose of Bioinformatics

What is Bioinformatics?

Bioinformatics against COVID-19

COVID origin

Protein structure prediction

Drug discovery

Funding

Datasets and resources

How to learn?

How to learn Bioinformatics?

COVID origin

Phylogenetic tree and BLAST

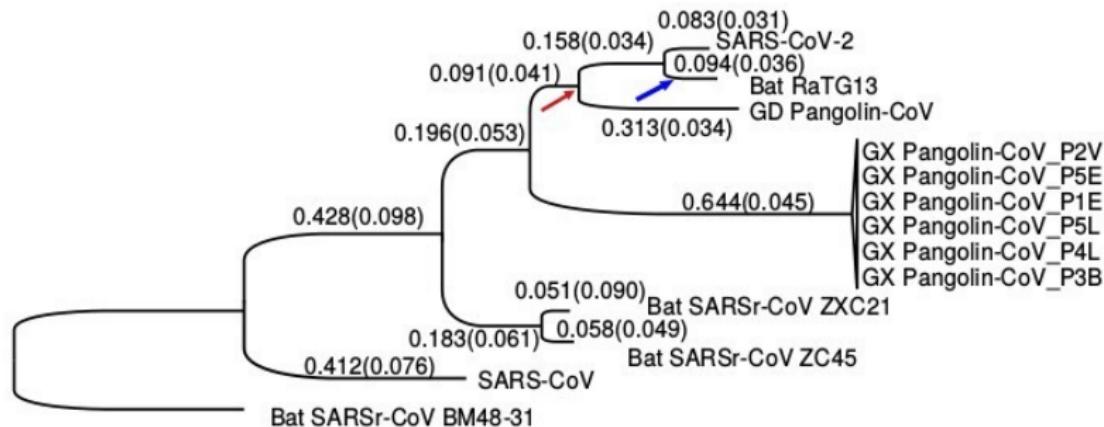


Figure: The phylogenetic tree of SARS-CoV-2 (COVID-19) and the related Coronaviruses [7].

COVID origin

Novel virus classification using alignment-free methods



22

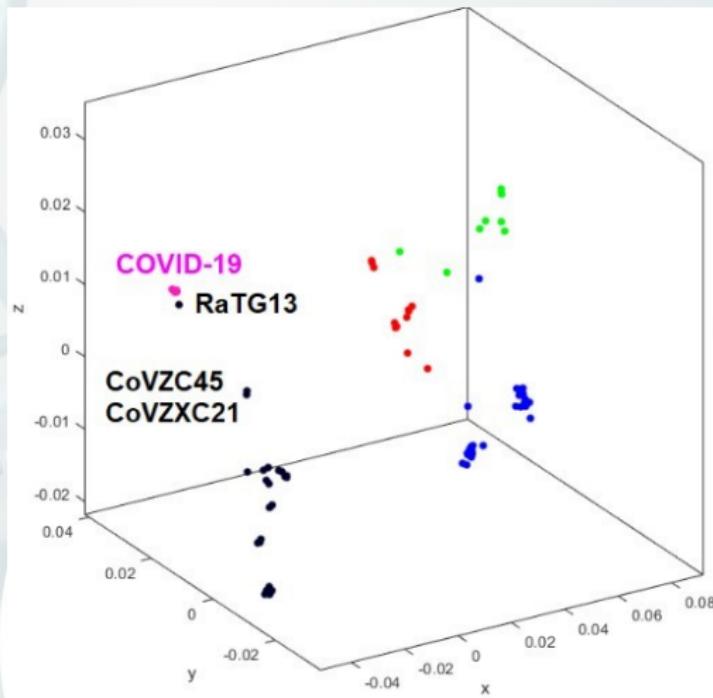


Figure: MoDMap3D of 124 Betacoronavirus sequences and COVID-19 [5].

Table of Contents



Introduction

Presentation

The purpose of Bioinformatics

What is Bioinformatics?

Bioinformatics against COVID-19

COVID origin

Protein structure prediction

Drug discovery

Funding

Datasets and resources

How to learn?

How to learn Bioinformatics?

Protein structure prediction

Definition



Definition

The prediction of protein three-dimensional structure from amino acid sequence [8].

Protein structure prediction

Definition



Definition

The prediction of protein three-dimensional structure from amino acid sequence [8].

Methods

- ▶ X-ray crystallography.
- ▶ Nuclear magnetic resonance.
- ▶ Cryo-electron microscopy.

Protein structure prediction

Using computers



There are two approaches to predicting protein structures:

- ▶ Homology modeling.
- ▶ Physical modeling.

Protein structure prediction

Proteins in COVID-19



Figure: Graphical view of COVID-19 structure. Source: [9]

Protein structure prediction

Proteins in COVID-19

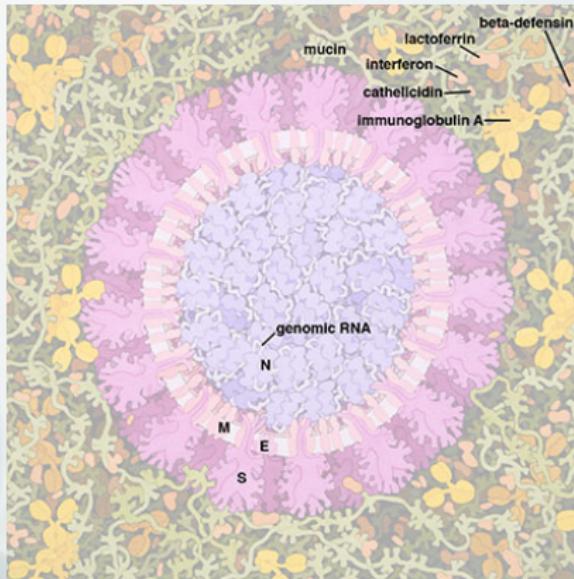


Figure: Membrane S (spike) protein, M (membrane) protein, membrane channel E (envelope) protein and the N (nucleocapsid) protein bound to the genomic RNA. Source: [9]

Protein structure prediction

AlphaFold method



28

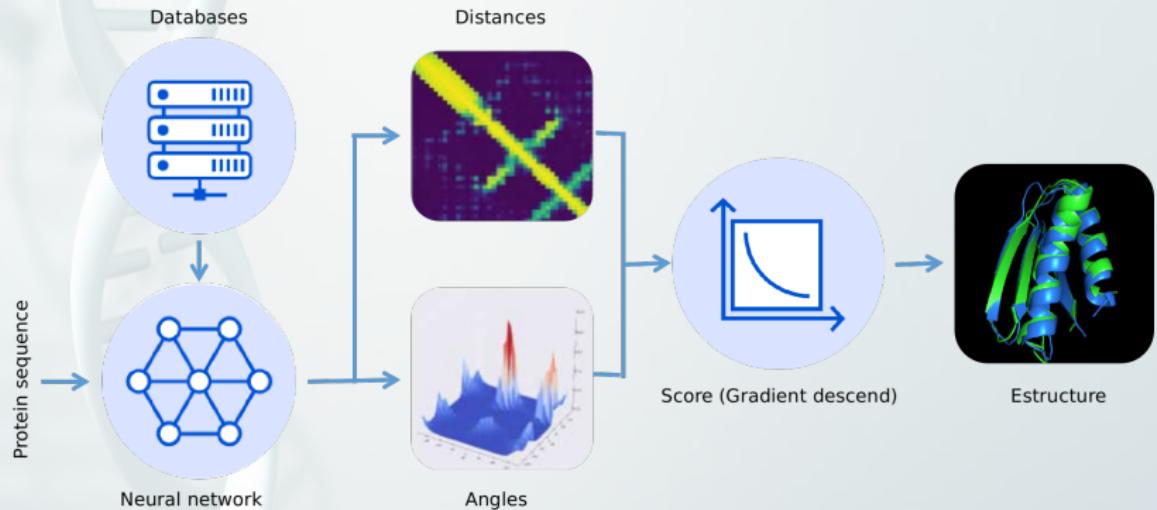


Figure: Protein structure prediction method proposed by AlphaFold. Source: [10]

Protein structure prediction

COVID-19 membrane protein

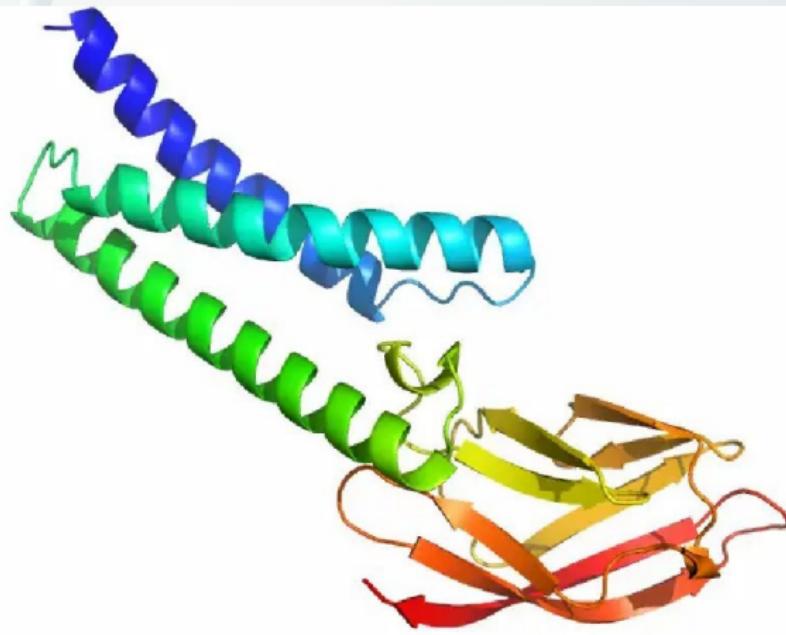


Figure: COVID-19 membrane protein. Source: [10]

Protein structure prediction

Tertiary representations of the S1 and S2 subunits of the spike protein

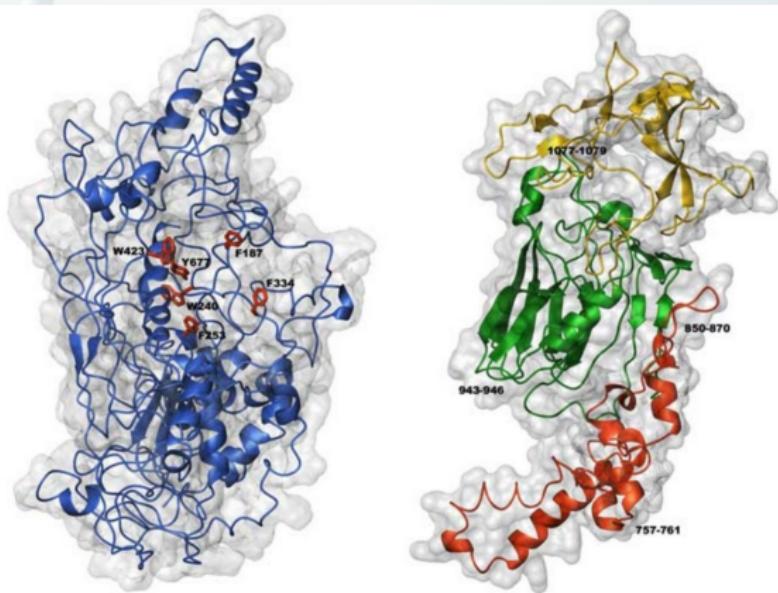


Figure: Tertiary representations of the S1 and S2 subunits of the spike protein using PsiPred. Source: [11]

Table of Contents



Introduction

Presentation

The purpose of Bioinformatics

What is Bioinformatics?

Bioinformatics against COVID-19

COVID origin

Protein structure prediction

Drug discovery

Funding

Datasets and resources

How to learn?

How to learn Bioinformatics?

Drug discovery

Definition



Definition

Drug discovery is the process by new candidate medications are discovered [12].



Molecular docking

Molecular docking is a computer simulation procedure to predict the conformation of a receptor-ligand complex [13]

Algorithms used:

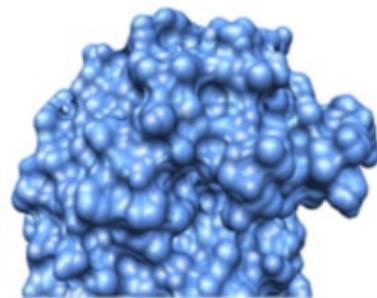
- ▶ Fast shape matching (take into account the geometric).
- ▶ Simulated Annealing.
- ▶ Genetic algorithms.
- ▶ Tabu search.

Drug discovery

Molecular docking



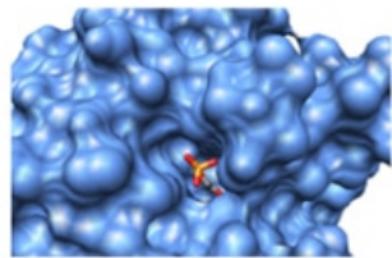
Target



Ligand



Molecular Docking

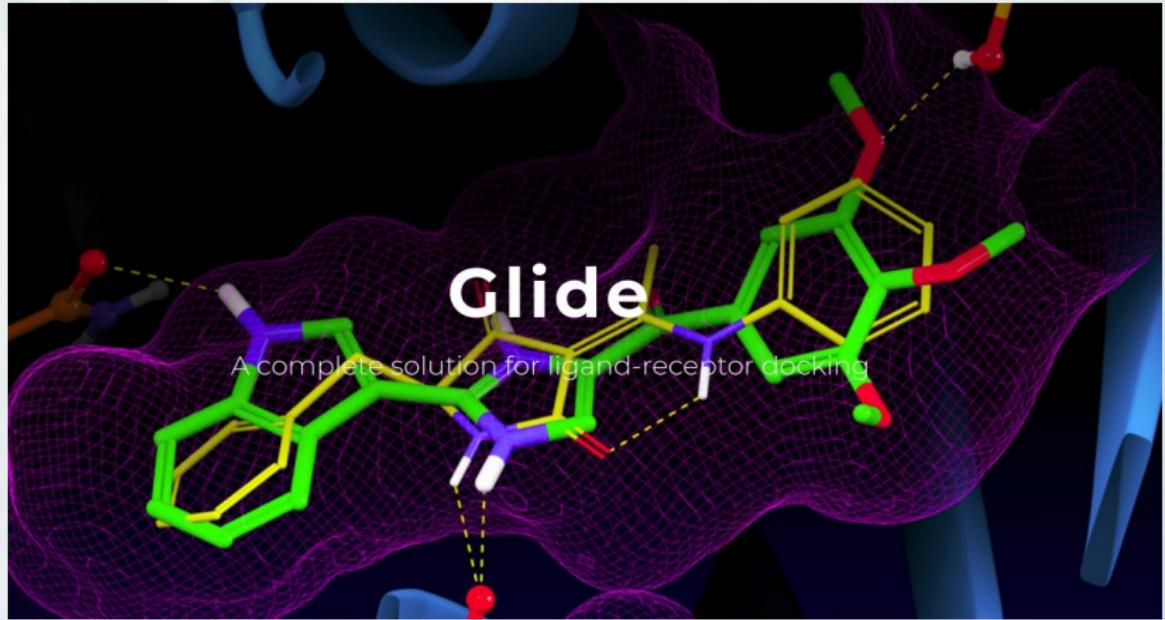


Drug discovery

Molecular docking with Glide



35



Drug discovery

From a million to one



36

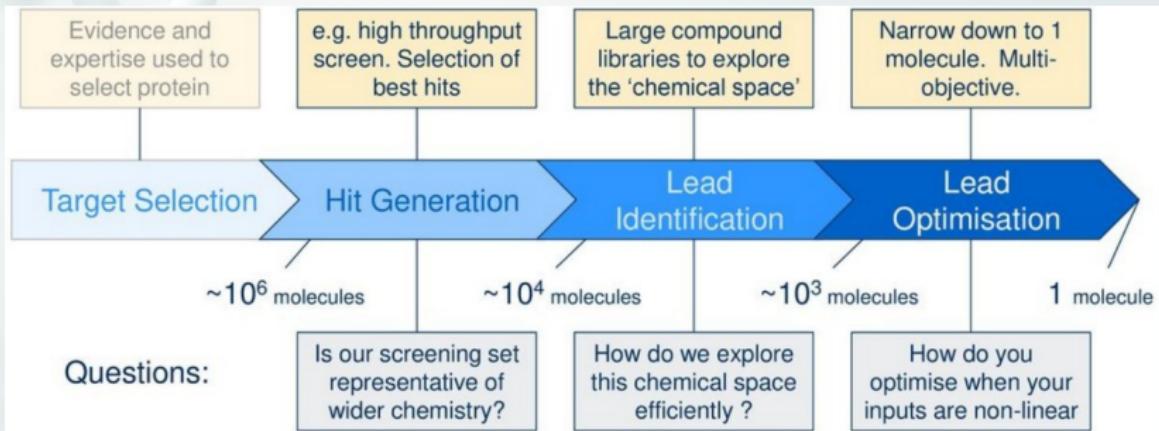


Figure: Process in drug discovery. Source: [14]

Drug discovery

COVID-19 main protease

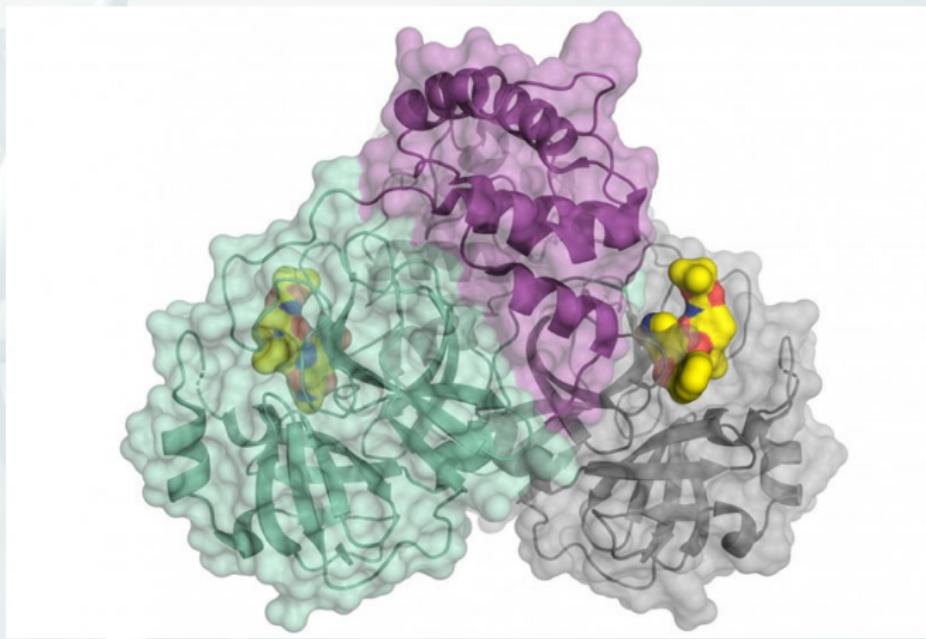


Figure: Schematic representation of the coronavirus protease. Source: [15]

Drug discovery

N3 inhibitor of main protease



38

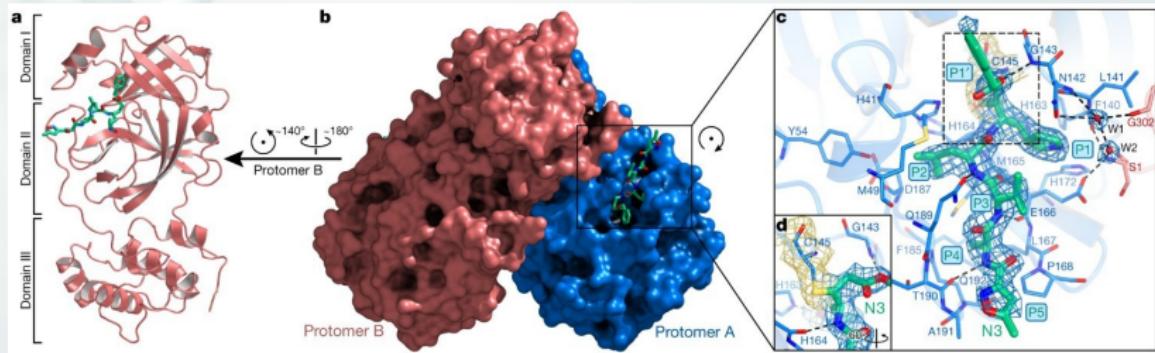


Figure: The crystal structure of SARS-CoV-2 main protease N3 inhibitor.
Source: [16]

Drug discovery

Protease Inhibitors Designed Using Generative Deep Learning Approaches



39

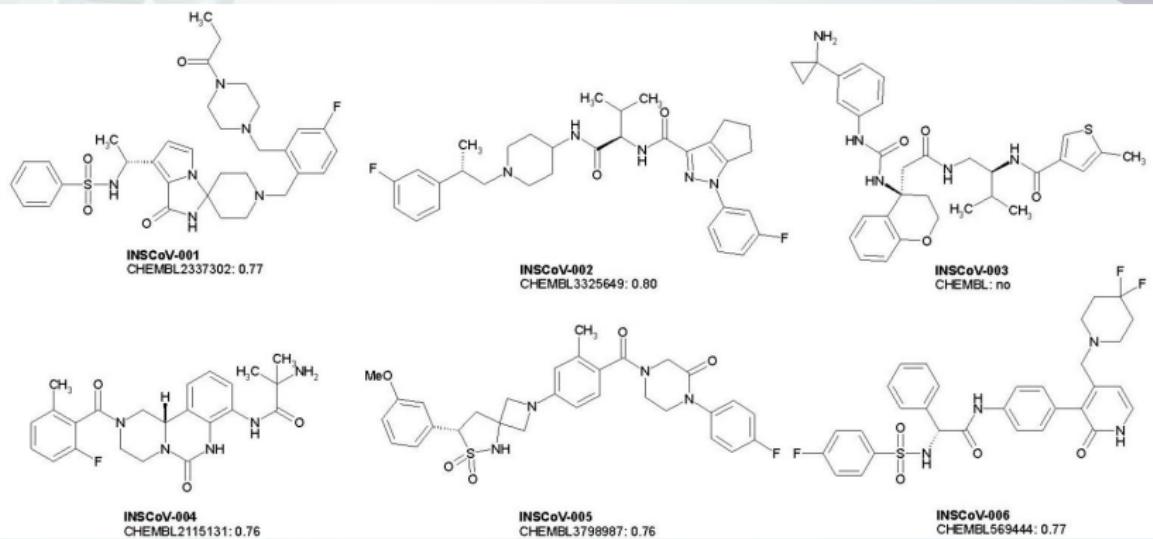


Figure: Representative examples of the structures generated to target the main protease of 2019-nCoV. Novelty was assessed using similarity search in ChEMBL Database. Source: [17]

Table of Contents



Introduction

Presentation

The purpose of Bioinformatics

What is Bioinformatics?

Bioinformatics against COVID-19

COVID origin

Protein structure prediction

Drug discovery

Funding

Datasets and resources

How to learn?

How to learn Bioinformatics?



List of open funding calls and other support for researchers, non-profit organizations and commercial organizations, specifically for COVID-19 and coronavirus-related research ([Link](#)).

Coronavirus Funding Monitor





MIT is hosting a series of challenges to empower YOU to take action on the COVID-19 crisis ([Link](#)).

MIT COVID19 CHALLENGE



FONDECYT

FONDO NACIONAL DE DESARROLLO CIENTÍFICO,
TECNOLÓGICO Y DE INNOVACIÓN TECNOLÓGICA

Table of Contents



Introduction

Presentation

The purpose of Bioinformatics

What is Bioinformatics?

Bioinformatics against COVID-19

COVID origin

Protein structure prediction

Drug discovery

Funding

Datasets and resources

How to learn?

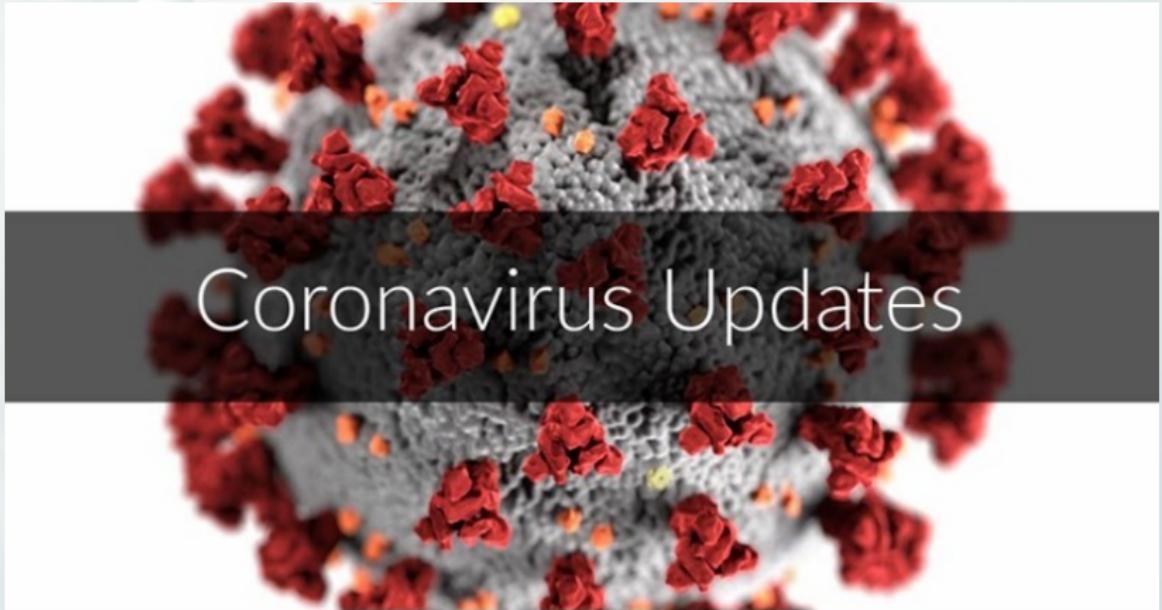
How to learn Bioinformatics?

Datasets and resources

Coronavirus updates



Links to bioinformatics resources useful to track the evolution and progression as well as to manage genomics data ([Link](#)).



Datasets and resources

Institut Français de Bioinformatique (IFB)



46

The IFB offers expertise and computing facilities to support the involved teams on COVID-19 ([Link](#)).



Datasets and resources

The European Bioinformatics Institute (EMBL-EBI)



47

EMBL-EBI is gathering and sharing data resources as they become available ([Link](#)).



Table of Contents



Introduction

Presentation

The purpose of Bioinformatics

What is Bioinformatics?

Bioinformatics against COVID-19

COVID origin

Protein structure prediction

Drug discovery

Funding

Datasets and resources

How to learn?

How to learn Bioinformatics?

How to learn Bioinformatics?

Biology as a DATA SCIENCE



49

T-Bio platform ([Link](#)).



How to learn Bioinformatics?

Coursera



Coursera ([Link](#)).

The Coursera logo, which consists of the word "coursera" in a lowercase, bold, sans-serif font. The letters are white and set against a solid blue rectangular background.

How to learn Bioinformatics?

Bioinformatics Research Group



51

Bioinformatics Research Group at la Salle university is a interdisciplinary group open for everybody who have a computer science' background.



References I



- [1] 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.
- [2] M. Clinics, "How genetic disorders are inherited," <https://www.mayoclinic.org/tests-procedures/genetic-testing/multimedia/genetic-disorders/sls-20076216?s=2>, 2020, accessed: 2020-03-20.
- [3] NAU, "Dna structure," http://www2.nau.edu/lrm22/lessons/dna_notes/dna_notes.html, 2020, accessed: 2020-03-20.
- [4] J. M. Archibald, *Genomics: A Very Short Introduction*. Oxford University Press, 2018, vol. 559.

References II



- [5] 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.
- [6] G. BIO, "Gen. bio,"
<https://sites.google.com/site/bio1040genbio2/home>, 2020,
accessed: 2020-03-20.
- [7] 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.
- [8] B. Kuhlman and P. Bradley, "Advances in protein structure prediction and design," *Nature Reviews Molecular Cell Biology*, vol. 20, no. 11, pp. 681–697, 2019.

References III



54

- [9] D. Goodsell, "Coronavirus," *RCSB Protein Data Bank*, Feb. 2020. [Online]. Available: https://doi.org/10.2210/rcsb_pdb/goodsell-gallery-019
- [10] A. Senior, J. Jumper, D. Hassabis, and P. Kohli, "AlphaFold: Using ai for scientific discovery," 2020.
- [11] O. Spiga, A. Bernini, A. Ciutti, S. Chiellini, N. Menciassi, F. Finetti, V. Causarono, F. Anselmi, F. Prischi, and N. Niccolai, "Molecular modelling of s1 and s2 subunits of sars coronavirus spike glycoprotein," *Biochemical and biophysical research communications*, vol. 310, no. 1, pp. 78–83, 2003.
- [12] U. Food and Drug, "The drug development process," 2020.
- [13] R. Dias, J. de Azevedo, and F. Walter, "Molecular docking algorithms," *Current drug targets*, vol. 9, no. 12, pp. 1040–1047, 2008.

References IV



- [14] C. Nantasesamat, "Computational drug discovery: Machine learning for making sense of big data in drug discovery," 2020.
- [15] L. Zhang, D. Lin, X. Sun, U. Curth, C. Drosten, L. Sauerhering, S. Becker, K. Rox, and R. Hilgenfeld, "Crystal structure of sars-cov-2 main protease provides a basis for design of improved α -ketoamide inhibitors," *Science*, vol. 368, no. 6489, pp. 409–412, 2020.
- [16] Z. Jin, X. Du, Y. Xu, Y. Deng, M. Liu, Y. Zhao, B. Zhang, X. Li, L. Zhang, C. Peng *et al.*, "Structure of m pro from sars-cov-2 and discovery of its inhibitors," *Nature*, pp. 1–5, 2020.

References V



- [17] A. Zhavoronkov, V. Aladinskiy, A. Zhebrak, B. Zagribelnyy, V. Terentiev, D. S. Bezrukov, D. Polykovskiy, R. Shayakhmetov, A. Filimonov, P. Orekhov, Y. Yan, O. Popova, Q. Vanhaelen, A. Aliper, and Y. Ivanenkov, "Potential COVID-2019 3c-like protease inhibitors designed using generative deep learning approaches," Feb. 2020. [Online]. Available: <https://doi.org/10.26434/chemrxiv.11829102.v2>



Thank you