International Journal of Family Medicine and Primary Care

Review Article
Published: 09 Apr, 2021

Applications of Bioinformatics Tools for Prevention COVID-19 Infection

*Ravi Teja KS[1] and Maya Datt Joshi[2]**

[1]Department of Biomedical and Bioinformatics Engineering, Shobhit Institute of Engineering and Technology (Deemed to be University), India

[2]Department of Biotechnology, Shobhit Institute of Engineering and Technology (Deemed to be University), India

Abstract

The large increase in number of coronaviruses has been found and also coronavirus genomes has been sequenced and found that the given an abnormal event to perform genomics studies and bioinformatics studies on the family of viruses. This coronavirus has large genome (26.4 kb to 31.7 kb) compare with among other RNA viruses. Coronavirus has large amount of G+C contents vary from 32% to 43%. Phylogenetically coronaviruses has 3 genera lineages namely as Alpha coronavirus, Beta coronavirus and Gamma coronavirus. So Beta coronavirus has again divided into 4 subdivisions namely as A, B, C and D. Coronaviruses are well established pathogens of animals and humans. Variable number of small amount of ORF presents between the conserved genes and also present in posterior of nucleocapsid gene. According to these applications of Bioinformatics studies has been involved to prevent the COVID-19. There is large amount of Bioinformatics NGS studies is going on for COVID-19 studies. Using Bioinformatics tools and software's we can predict the protein structure and also identify which genes are responsible for causing coronavirus.

Introduction

Coronaviruses belong to the family *coronaviridae* in the *Nidovirales* order coronaviruses are minute in size (65 nm to 25 nm) and contain the single stranded RNA as nucleic material, size ranging from 26 kbs to 36 kbs in length. The life cycle of SARS-CoV-2 in host cells Every life cycle involves in three steps there are Attachment to the host cell receptor (Ig like receptors, cellular adhesion, molecules, membrane transport proteins, oligosaccharides etc.). Penetration (endocytosis, fusion) Uncoating Similarly, COVID-19 begins its lifecycle when the S-protein binds to the cellular receptor ACE2 After receptor binding, the conformation change in the S protein facilitates viral envelope fusion with the cellular membrane through the endosomal pathway The SARS-CoV-2 releases RNA into the host cell Genome RNA is translated into viral replicase polyproteins ppl a and lab, which are then cleaved into small products by viral proteinases. The polymerase produces a series of subgenomic mRNAs by discontinues transcription and finally translated into relevant proteins viral proteins and genome RNA are sub sequently assembled into virions in the ER and Golgi and then transported via vesicles and released out of the cell The entire mechanism of pathogenicity of SARS-CoV-2, from attachment to replication.

Tools and software

There are many tools involved in these studies.

BLAST (Basic Local Alignment Search Tool): It is a set of programs designed for the windows platform and it is used for similarity searches for protein or DNA sequence.

It is used to compare the novel sequence of protein or nucleotide sequence with template sequence or exit sequence i.e. previously characterized gene in the database.

Using these BLAST tool we can work on novel COVID-19. Then it will decide that COVID-19 is near to which virus or near or belongs what family.

GLIMMER: it means Gene Locator and Interpolated Markov ModelER. This is the first bioinformatics tool to identify genes. While retrieving result the accuracy of the result is approximately 96% to 98%. This software is available in TIGR (The Institute of Genome Research).

Trans term: Is the database providing access to mRNA sequences associated with the regulatory elements.

OPEN ACCESS

Correspondence:
Maya Datt Joshi, Department of Biotechnology, Shobhit Institute of Engineering and Technology (Deemed to be University) Modipuram Meerut, India,
E-mail: mayajoshi2778@gmail.com

Received Date: *17 Mar 2021*
Accepted Date: *06 Apr 2021*
Published Date: *09 Apr 2021*

Citation:
Ravi Teja KS, Joshi MD. Applications of Bioinformatics Tools for Prevention COVID-19 Infection. Int J Fam Med Prim Care. 2021; 2(2): 1034.

Tandem repeat finder: Is the database software tool used for identifying the tandem repeats in the sequence.

Tandem repeats which means consists of two or more than approximate nucleotides placed very close to each other in the DNA sequence.

RBS finder: It is the databases of software tool which can be used identify the ribosome binding site in the mRNA regions.

RBS is a sequence in mRNA which can help to initiate translation.

Swiss modeler: is an online tool which can be used for to construct protein structure based on the sequence and initially identify the templates and then build the model then we get complete protein structure. Then we will use this protein structure for further studies. Coming COVID-19 studies we will use these protein structure for docking studies and further for protein interactions studies.

Rasmol: Is a computer program which can be used for visualize the protein structure from Protein Data Bank (PDB). It is a scientific software tool and open free software. Which this software helps to make the protein structure in different forms like ribbons, loops, wires etc. It means make own forms based on our interest. Based on this software we can see the protein structure in 3D visualization. Which can help to identify the amino acid regions for protein interactions and drug interaction studies?

DAVID (Database for Annotation Visualization and Integrated Discovery): Which these DAVID tools identify based on GEO (Gene Expression Omni Bus) which used for functional classification and gene classification studies. This DAVID is completely based on gene set of data and list of genes, if we submit the gene list of the organism or species. Then it will give us possible pathways of our gene list of data. The pathways like KEGG, Biocarta etc. pathways involved in these studies.

We use these type of studies in COVID-19 these gene list will retrieve from the GEO of NCBI. This DAVID most probably used for RNA-seq data analysis studies. So initially we get the SARS-COV-2 GEO data from NCBI and run program studies using Perl program. Which these Perl program studies make to differentiate the data based on p-values. By this way we get complete gene list of coronavirus then we will get pathways then we see the pathways and make a graph according to the cutoff value.

MEGA (Molecular Evolution Genetic Analysis): Is nothing but molecular evolutionary genetic analysis. It is computer software it is used constructing statically analysis of constructing molecular evolution and phylogenetic tree. There are many tools for phylogenetic tree and phylomedicine. It is free software accessible to every user.

UCSF chimera: Is a program for interactive visualization and analysis of molecular structures including density maps. Trajectories and sequence alignment it is open free software. It is also of 3D visualization and interactions of molecular structures. And based on software we can select different types of modules in software like change the color and chemistry according to our needs. In this software we have tools which can work in different and accuracy in this which we have surface and binding analysis and plane and geometric dimensions to design protein and also see the stability of our viral protein.

According to COVID-19 studies we can have a possible to do docking and protein interaction studies of viral protein and drug studies.

IEDB (Immunological Epitope Data Base): is a free resource maintained by National Institute of Allergy and infectious Diseases. It is used for searching and experiment of characterizing antibody and T-cell epitopes studied and humans and non-humans primates.

Gene mania: It is an online tool which we can see the see the what type of genes are interacted with to our gene (target gene) and what type of interaction like physical interaction, genetic interactions and gene expression levels and co localization and also predicted levels of genes so based on these type of tool which will get the complete gene list and we will go for further studies.

So, based on gene mania tool we can also predict what type of gene interactions, predicted domains will happen in coronavirus and also show coronavirus gene interactions will be either strong or weak and also know what type of genes interacted with coronavirus and that genes are available in which organs and also we predict that these corona viral genes will cause any damage our body or can mutate our metabolism [1,2].

String: It is an online tool which can say that which genes are interacting with proteins for protein studies. It is completely protein based tool .Interaction of each and every protein will be in different colors for easy understanding. If we want to know which protein are involved in your target protein then this tool is very helpful to tell you the protein interaction studies. This string is a protein association network. In coronavirus studies it will help the proteins are involved or interaction with corona viral protein. It will also say the respective protein belongs to which family and also tell that particular protein from which organs like liver, lungs kidneys etc. [3-5].

FIMM: The database by Schoenbach & Brusic [6] is a functional database of molecular immunology. The database contains 571 antigens and 1,591 peptides.

MHCBN: Bhasin [7] is a database of MHC binding and non-binding peptides containing 14,816 binders, 1,782 non-binders and 5,456 T-cell epitope entries.

HLA Ligand/Motif database: This site's database can be searched by defining allele and specificity, amino acid pattern, ligand/motif in sequence of amino acids, author's last name, or advanced search with more criteria.

HIV Molecular Immunology database: The HIV Molecular Immunology Database is an annotated, searchable collection of HIV-1 cytotoxic and helper T-cell epitopes and antibody binding sites. The goal of the database is to provide a comprehensive listing of defined HIV epitopes [8].

PAProC: Prediction algorithm for proteasomal cleavages is a prediction tool for cleavages by human and yeast proteasomes, based on experimental cleavage data [9]. An updated version of the PAProC program based on *in vitro* immunoproteasome cleavage data is also in the making according to the PAProC homepage [10].

IMGT: IMGT, the international Immunogenetics project, is a collection of databases specializing in Immunoglobulins, T cell receptors and the Major Histocompatibility Complex (MHC) of all vertebrate species. The IMGT project was established in 1989 by the University Montpellier II and the CNRS (Montpellier, France) and works in close collaboration with the EBI [11-14].

ASHI: The American Society for Histocompatibility and

Immunogenetics (ASHI) hosts databases of gene and allele frequencies (www.ashi-hla.org/).

Predict: The Predict method use neural networks to predict Class I, II and TAP binding.

BIMAS: The BIMAS method was developed by Parker. The method is based on coefficient tables deduced from the published literature. For HLA-A2, peptide binding data were combined together to generate a table containing 180 coefficients (20 amino acids x 9 positions), each of which represents the contribution of one particular amino acid residue at a specified position within the peptide.

Conclusion

Based on all these type of Software's and tools we can perform our coronavirus studies easily and freely and these tools which can make our work easily and will give accuracy results. And finally based on these software and tools we can predict few outputs of coronavirus which can get preventive measures. Few of the vaccine databases will give few information or prediction of coronavirus studies. By using these tools and software's we just make to preventive or tell in research point of view. We will know what exactly happening in genetic studies and protein studies and metabolic studies of coronavirus. Initially we get primary information from wet lab studies. So taking primary information as template we will further proceed to secondary information as dry lab i.e. insilico studies. These are the main studies which can be done by software, tools and programming studies. Finally I conclude that this is the one of the way to know the genetic, protein studies etc to know or I predict to get an idea to prevent coronavirus.

Discussion

Based on above information and tools we can say that what actually happening in the coronavirus genes and protein studies we can say that these studies tell us which genes and which chromosome cause the virus. And also possible to do comparative studies it can be done by using Perl programming studies it shows us what is the expression of one gene compare with other gene based on cut off values and also perform statistic studies then we will conclude that we will clear picture that which gene is highly expressed and which genes are low expressing genes.

References

1. Patrick Woo CY, Huang Yi, Susanna Lau KP, Yuen KY. Coronavirus genomics and bioinformatics analysis. Viruses. 2010;2(8):1804–20.
2. Susan Payne Department of Veterinary Medicine and Biomedical Sciences, Texas A&M University, College Station, Texas, United States.
3. Goodswen Stephen J, Kennedy PJ, Ellis JT. Evaluating high-throughput *Ab Initio* gene finders to discover proteins encoded in eukaryotic pathogen genomes missed by laboratory techniques. Plos One. 2012.
4. Wegmann U, Horn N, Cardinga SR. Defining the bacteroides ribosomal binding site. 2013;79(6):1980-89.
5. Kumar S, Nei M, Dudley J, Tamura K. MEGA: A biologist-centric software for evolutionary analysis of DNA and protein sequences. 2008;9(4):299-306.
6. Schönbach C, Koh JLY, Flower DR, Wong L, Brusic V. FIMM, a database of functional molecular immunology: update 2002. Nucleic Acids Res. 2002;30(1):226-9.
7. Bhasin M, Singh H, Raghava GPS. MHCBN: A comprehensive database of MHC binding and non-binding peptides. Bioinformatics. 2003;19(5):665-6.
8. Yusim K, Addo MM, Brunak S, Chigaev A, Detours V, Korber BT, et al. Clustering patterns of cytotoxic T-lymphocyte epitopes in Human Immunodeficiency Virus type 1 (HIV-1) proteins reveal imprints of immune evasion on HIV-1 global variation. J Virol. 2002;76(17):8757-68.
9. Nussbaum AK, Kuttler C, Hadeler KP, Rammensee HG, Schild H. PA ProC: A prediction algorithm for proteasomal cleavages. Immunogenetics. 2001;53(2):87-94.
10. Lefranc MP. IMGT, the international ImMunoGeneTics database. Nucleic Acids Res. 1999;27(1):209-12.
11. Lundegaard C, Lund O, Nielsen M. Prediction of epitopes using neural network based methods. J Immunol Methods. 2011;374(1-2):26–34.
12. Lund O, Nielsen M, Lundegaard C, Kesmir C, Brunak S. Immunological Bioinformatics. USA; MIT Press: 2005.
13. Kritikos M. Ten technologies to fight coronavirus in-depth analysis EPRS | European Parliamentary Research Service. Scientific Foresight Unit (STOA) PE 641.543 – April 2020.
14. Mark R, Denison MD. Coronavirus research: Keys to diagnosis, treatment, and prevention of SARS. In: The National academies forum on microbial threats board on global health. Washington Dc; National Academic press: 2004.