Influenza Vaccine Prediction Based On Viral Strains

A Systematic Literature Review

Team Name : Go Corona

Contributors

Anmol Kumar
Samarth Chauhan
Ritvik Gupta
Karan Chawla
Anjali Sharma

Abstract

Influenza remains to be one of the biggest viral infections and has a huge impact on global health every year. With the constant evolution of the viral strains and the mortality and morbidity rate of the viruses, it becomes highly necessary that we refine our methods for the effective drug production based medical research. In this systematic review, we have conducted an online search about the response of different viral domains considering their mutating/ non-mutating characteristics. We try to address the development and challenges of the universal influenza vaccine. Based on the search, we introduce our machine learning prediction model based on the SNP sites where mutation/antigenic drift is frequent.

Introduction

Influenza, commonly known as the flu, is caused by a **virus** that attacks the upper respiratory tract (i.e., the nose, the throat and the lungs). Cold and dry weather allows the virus to survive longer outside the body than in warm weather.

There are three **types of influenza virus**: A, B and C. Type A can infect humans, other mammals and birds and can spread fast and affect many people. Types B and C affect only humans and type C causes only a mild infection. Influenza type A viruses are sub-typed into two categories based on proteins, specifically the proteins *hemagglutinin* and *neuraminidase*, on the surface of the virus. The virus uses the hemagglutinin protein (often abbreviated "H" or "HA") to latch on to the host's cell and uses the neuraminidase protein (often abbreviated "N" or "NA") to spread the infection. Types A and B viruses continually evolve genetically, with changes being made to the *amino acid sequence* of the H and N proteins. Since hosts recognize the H and N **surface proteins** to identify and attack the virus, by changing these proteins a little bit the virus prevents the hosts from enjoying any prolonged protection against the virus.

When a person is vaccinated with the influenza **vaccine**, it should stimulate a protective immune response, particularly against the viral surface proteins in the viral strains used to make the specific vaccine. The influenza vaccine typically contains three **virus strains**, two are subtypes of type A and one is of type B. Type C is not included in the vaccine because it only causes a mild illness and does not lead to **epidemics**. To make the influenza vaccine, gene fragments that encode the H and N viral surface proteins are used from each strain. For the vaccine to give a person good protection against the virus, the **protein sequences** for the H and N proteins that are used in the vaccine should closely match the sequences in the strains the person may be exposed to.

General questions

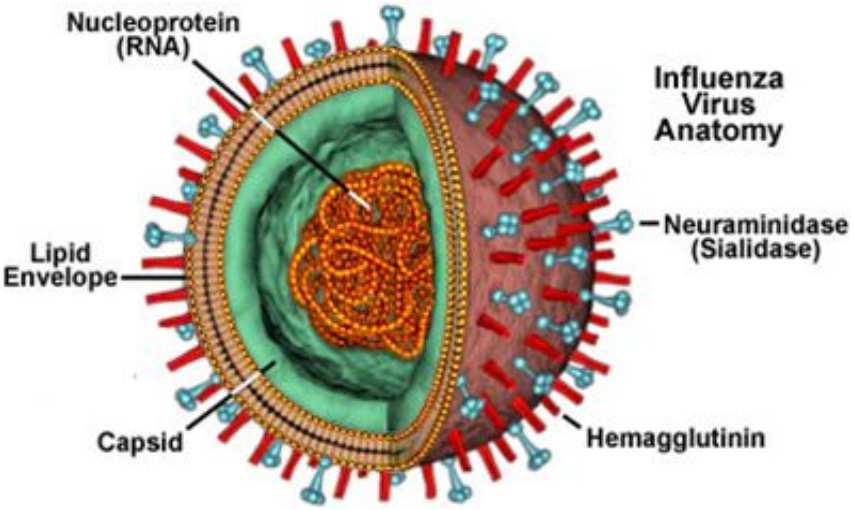
- How are flu viruses named ?
- BLAST algorithm can be used to align and compare both DNA (nucleotide) sequences and protein (amino acid) sequences. What are some reasons for using a protein alignment instead of a DNA alignment ?
- There are many H and N subtypes for influenza A. Why is it that in recent years the annual vaccine has only included influenza A subtypes H1N1, H3N2 and B type virus? What is happening with the other subtypes? Under what conditions might they be included in the annual vaccine ?
- How does a vaccine help prevent the spread of a disease ?
- Can vaccines help if a person already has the virus ?
- Does antigenic drift have a pattern ?
- Can we make a universal influenza vaccine ?

Background

First we will start with what flu is. Flu is a contagious respiratory illness caused by influenza viruses that infect the nose, throat, and sometimes the lungs. It can cause mild to severe illness, and at times can lead to death.[1] Influenza illness which is caused by influenza virus is very trivial nowadays thanks to research & studies done in this area and vaccines. Alone in the US around 45 million illnesses were reported but only 0.015 million deaths were reported by influenza virus. So, around 99.9% lives were saved by influenza vaccine. Influenza virus (swine H1N1) was pandemic in the year 1918 which took around 350 million lives [2] as influenza virus can be easily spread through cough and respiratory tract.

Influenza viruses are rather variable in size and morphology. After initial isolation, they frequently occur as filaments, but after adaptation to the chick embryo, the viruses usually appear as spheres of about 100 nm in diameter. Purified preparations of influenza virus contain protein (60-75%), lipid (18-37%), non-nucleic acid carbohydrate (5-7%), and ribonucleic acid (0.8-1.0%).

Influenza virus uses our cell nucleus, ribosome to duplicate themselves until the cell dies.There are three types of influenza viruses, A, B, and C; these divisions are based on immunological differences between the internal matrix and nucleoproteins. A & B type virus contain 8 types of RNA while C type contain 7 types of RNA [3]



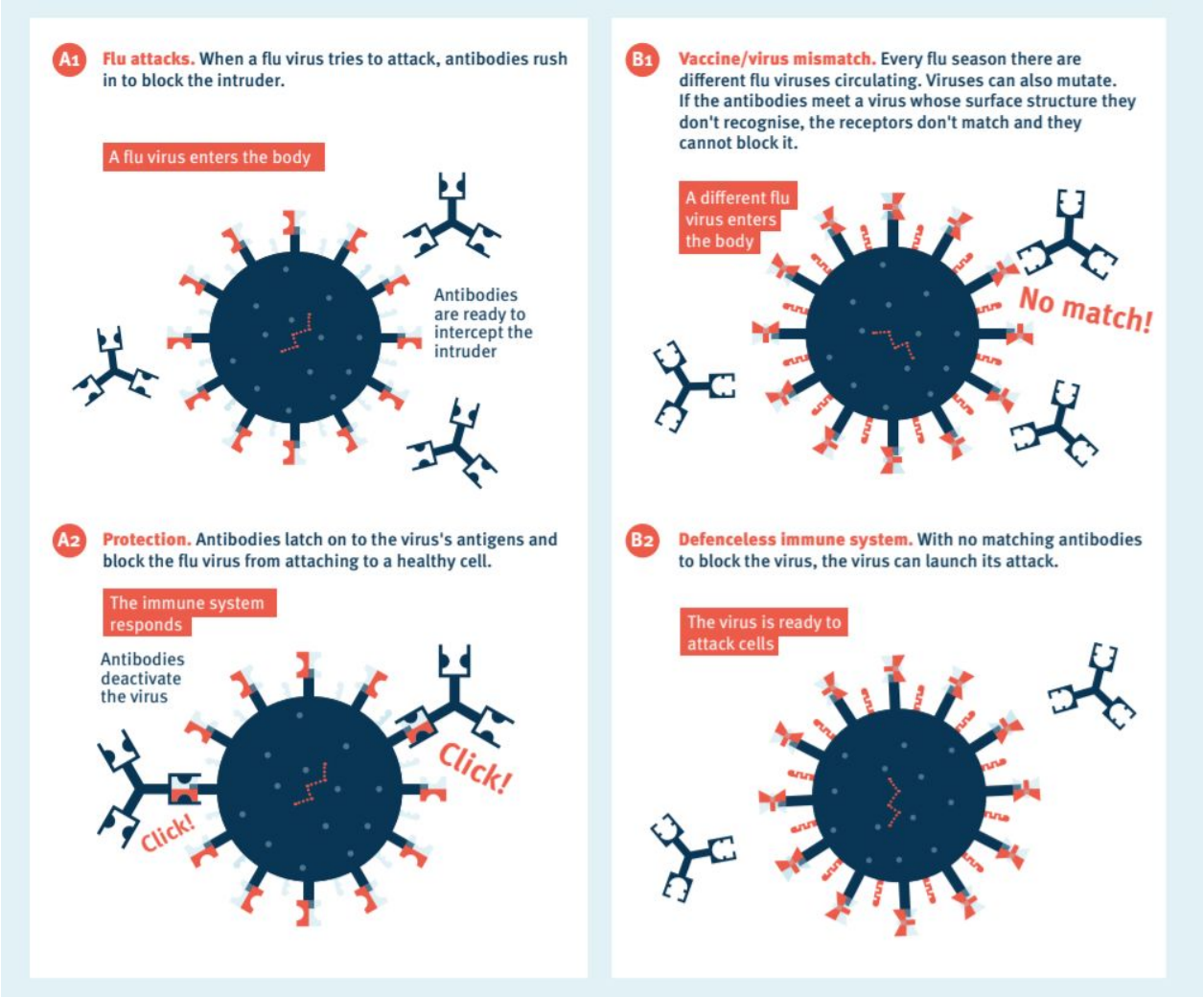
[Source]

Type	Symptoms	Epidemic	Vaccine	Genetic drift	Genetic shift	Pandemic	Animals Affects
A	Respiratory / Constitutional	Yes	Yes	Yes	Yes	Yes	Birds, Pigs, Horse, Dogs
B	Respiratory / Constitutional	Yes	Yes	Yes	No	No	No
C	Respiratory	No	No	Yes	No	No	Pigs, Dogs

Most of the vaccines target Neuraminidase (NA) and Hemagglutinin (HA) protein of the virus. HA protein on viruses helps viruses to attach to the receptor on the cell which after it can get into the cell and use its resources. Predicting area or SNPs site where mutation is frequent and find possible combinations of viruses. Our body produces antibodies in response to these viruses which are attached to HA protein like cover hence make it unable to attach it to cells.
Vaccines have viral strains in inactive or mildly active form, which are injected either by injection or nasal spray into the bloodstream and then our body creates antibodies against the viruses. So if the body is attacked by a similar virus in the future, our body can fight with it easily.

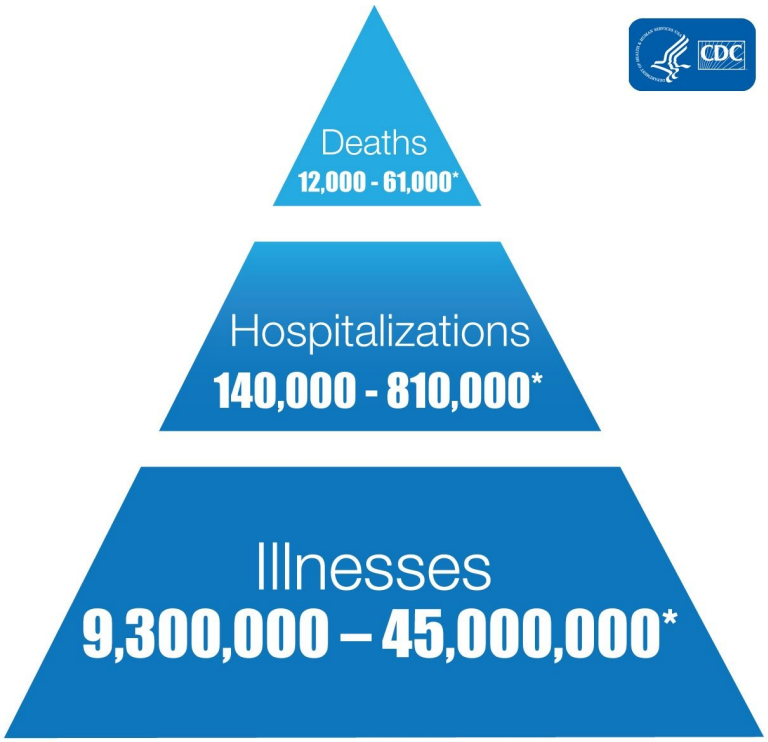
The influenza viruses in the seasonal flu vaccine are selected each year based on surveillance data indicating which viruses are circulating and forecasts about which viruses are the most likely to circulate during the coming season. The degree of similarity between available vaccine viruses and circulating viruses also is important. Vaccine viruses must be similar to the influenza viruses predicted to circulate most commonly during the upcoming season. Another important practical factor in the recommendation about what viruses to include in a flu vaccine is whether or not there is a good vaccine virus available; that is, a virus that could be used in vaccine production and which would likely protect against the viruses likely to circulate during the upcoming season.

Vaccine viruses were required by FDA to be isolated and grown in chicken eggs, but now the FDA allows vaccine viruses to be grown in cells, too. Regardless of how they are grown, vaccine viruses must be tested and available in time to allow for the production of the large amount of vaccine virus needed to make vaccines. Occasionally, a suitable vaccine virus cannot be identified or developed in time to be included in the upcoming season's vaccine.



[Source]

The greatest risk remains with young children below the age of 5 and particularly those under 2, pregnant women, elder adults over the age of 65 and residents of nursing homes and hospitals. Children and elder people generally have higher risk as they have reduced immunity resistance to viruses. Every year, Centers for Disease Control and Prevention (CDC) estimates the number of influenza cases and its impact on the population. It also reports the impact of the influenza vaccines on the population. The official CDC website states that influenza has resulted in between 9 million – 45 million illnesses, between 140,000 – 810,000 hospitalizations and between 12,000 – 61,000 deaths annually since 2010.



[Source]

The need of a ‘Universal’ Influenza vaccine

With all the research and resources that go into the development of seasonal influenza vaccines annually, it becomes highly necessary to develop a universal influenza vaccine. But what do we exactly mean by this? Universal influenza virus vaccines are vaccines that cover all

influenza A and influenza B viruses independent of antigenic drift or HA/neuraminidase (NA) subtype. This would include all 18 HA and 11 NA subtypes of influenza A.

Thus we need to look at the targeted responses by the vaccines which are independent of HA/NA subtype responses. This could be achieved by a vaccine strain which is based on a conserved part of the influenza virus and is unaffected by antigenic change which occurs in the viruses throughout the year. The HA and NA subtype are highly antigenic parts of the virus. But research has found that there is the involvement of other extracellular domains like the ectodomain of the M2 ion channel protein and the internal proteins nucleoprotein(NP) as well. Both are considered as conserved antigens as they have low rates of mutation and thus these domains would be highly favourable sites for universal vaccine development. [5][6]

The dependence of the functioning of the current vaccines dependent on antigenic drift or HA/neuraminidase (NA) subtype responses is the reason that we are not able to combat this pandemic. There are many expectations from this universal influenza vaccine such as it will entirely replace the seasonal influenza vaccine that are currently developed annually and are then circulated. But there are still many challenges, first and foremost being that several plausible universal vaccine solutions are still in a preclinical/clinical developmental stage. Therefore, it is still too early to expect anything. These plausible candidates can face economic as well as regulatory challenges, the cost of trials and approval at a large scale. Even if the universal vaccine is made, another challenge which arises is maintaining the persistence of the effect of the vaccine and the coverage of diverse range of human populations with the use of **adjuvants** (to induce a broader response within the body). A vaccine which would last for a shorter period of time is of very limited importance. It would truly be ‘universal’ if any race of people, living in any part of the world, could get the desired response from it.

We are still very far before we could achieve a universally approved vaccine that would control the development and spread of the seasonal influenza. But meanwhile, newer technologies and methods for early prediction of the possible viral strains for the upcoming years are required or are needed to be refined so that the vaccines could be made way earlier than the current production period. For this we will be trying to work upon prediction of patterns among the antigenic sites of the viral strains of the influenza viruses.

Our project model

In this project, we are trying to map the best vaccine available or best strain from which vaccine stain can be made by using virus and vaccine strain of the previous years from 1972 to 2020. For vaccines, we will take virus strain sequence from the victim's body and will try to find similar or identical virus strain in the database. If a match is found, then vaccine stain for the same virus can be given to the victim and then we will also show a prediction confidence/ probability number to vaccine strain for this virus by which it can treat or cure the influenza virus illness.

When a virus replicates in the cell it can get mutated by recombination, insertion or deletion, also called antigenic drift. Antigenic drift can occur between the same virus or between various viruses as they can replicate in the animals also. Most common influenza viruses which attack humans are A & B type viruses. There are 18 types of HA protein and 11 types of NA protein in A type virus. H1N1 and H3N2 affect most in the A type virus and these can be pandemic. Our database will contain a sequence of virus and strain of H1N1 (A type), H3N2 (A type) and B type of virus and their mutation version throughout the years. List of the virus strain and vaccine is provided in the fludb website while we will take sequence NCBI Genbank. We will store viral strain and its corresponding vaccine strain and then we will match the sequence of victims viral strain to the one present in our database.

For the prediction part we are trying to find patterns in the sequence of the virus strain which can be done by mutation i.e. antigenic drift or possible combination which can come if 2 or more viruses interact. To find patterns, we will use previous year sequence and by machine learning we will find the prediction sequence.

Predicting SNP sites where mutation is frequent and find possible combinations of virus

There are models in machine learning that can find sites or areas where mutation is frequent and chance of getting a virus mutated at that place. These models can help government and pharma companies to make vaccines in advance and this also leads us to the path of universal vaccine because we know all possible combinations of viruses that can affect the population. For this we will consider sequences of only particular influenza viruses like H1N1 and then we will try to find similarities between different sequences that are available in our database by phylogenetic tree or by clustering etc. [4]

Is a similar study done earlier or not?

No specific study or systematic review has been in the manner we are doing. But there are a few reviews that have partially similar questions that we are addressing.
There are studies focused predicting vaccine strains for a particular virus only. In contrast, Our methodology is different and we are considering a large variety of virus strains, mainly H1N1 A type, H3N2 A type and B type viruses.

Here are some of the related reviews:

1) **The determinants of 2009 pandemic A/H1N1 influenza vaccination: a systematic review.**
   https://www.sciencedirect.com/science/article/pii/S0264410X1102038X?via%3Dihub

2) **Models for Predicting the Evolution of Influenza to Inform Vaccine Strain Selection:**
   https://www.ncbi.nlm.nih.gov/pmc/articles/PMC5861780

Research questions

Following questions are addressed in our paper:

1. How to predict which viral strain must be used for making vaccines for any new virus?
2. To predict the sites of SNP on particular flu viruses like H1N1/H3N2 and make vaccines in advance.
3. How can we move towards a Universal Influenza Vaccine?

Outcomes / Categories of information desired

1. To give the name of the virus strain which can be used to make a vaccine if it is similar to previous to previous virus strain.
2. To find patterns in the sequence of the virus strain to make a vaccine in advance.
3. To look for ways which can lead to an Universal Influenza Vaccine

Inclusion / Exclusion Criteria for your review

1. We studied only H1N1 (A type), H3N2 (A type) and B type virus and their behaviour as humans mostly affected by this virus strain.
2. Pathway analysis of viruses into cells is not considered in this review.
3. Database and study related to the USA is considered for this review as high quality data and research is done and available for the USA region but list includes samples from other countries also which was studied in the USA.

4. Age and immune system of the victim is not considered in the prediction of the vaccine.
5. Virus sequence data is downloaded from NCBI genbank and fludb only as it contains high quality annotated data.

Borderline cases

1. There were many sites which provided lists of virus and vaccine strain list but we only shortlisted those samples which are approved or made by WHO as they have the highest impact and reputed data related to it.
2. We only considered and included those papers which have an impact factor greater than 3. There were either those journals that have either This impact factor is provided by the google search engine.
3. There were studies focused on predicting vaccine strains for a particular virus only.

Search Strategy and Inclusion/Exclusion of Studies

We started searching from basic terms and got a large number of results. From there, We kept improving and refining our search to get only the relevant results in the direction of the aim of our systematic review. Once we reached a good level of refinement(search 7), we started screening out the resulting studies and included/excluded them based on their relevance.

The following is the search term corpus we have created.

Search 1: Influenza Virus 75,321 results

Keywords, synonyms:
(("orthomyxoviridae"[MeSH Terms] OR "orthomyxoviridae"[All Fields]) OR ("influenza"[All Fields] AND "virus"[All Fields])) OR "influenza virus"[All Fields]

Search 2: Influenza Virus Types 333 results

Keywords, synonyms:
((("orthomyxoviridae"[MeSH Terms] OR "orthomyxoviridae"[All Fields]) OR ("influenza"[All Fields] AND "virus"[All Fields])) OR "influenza virus"[All Fields]) AND "types"[All Fields]

Search 3: Influenza Virus Sequence 519 results

Keywords, synonyms:
(((("orthomyxoviridae"[MeSH Terms] OR "orthomyxoviridae"[All Fields]) OR ("influenza"[All Fields] AND "virus"[All Fields])) OR "influenza virus"[All Fields]) AND ((((((((((((((("base sequence"[MeSH Terms] OR ("base"[All Fields] AND "sequence"[All Fields])) OR "base sequence"[All Fields]) OR "sequence"[All Fields]) OR "sequences"[All Fields]) OR "sequence analysis"[MeSH Terms]) OR ("sequence"[All Fields] AND "analysis"[All Fields])) OR "sequence analysis"[All Fields]) OR "sequencing"[All Fields]) OR "sequence s"[All Fields]) OR "sequenceable"[All Fields]) OR "sequenced"[All Fields]) OR "sequencing"[All Fields]) OR "sequencer"[All Fields]) OR "sequencers"[All Fields]) OR "sequences"[All Fields]) OR "sequencings"[All Fields])

Search 4: Influenza Virus Vaccine **6481 results**

Keywords, synonyms:
(((("influenza vaccines"[MeSH Terms] OR ("influenza"[All Fields] AND "vaccines"[All Fields])) OR "influenza vaccines"[All Fields]) OR (("influenza"[All Fields] AND "virus"[All Fields]) AND "vaccine"[All Fields])) OR "influenza virus vaccine"[All Fields]
Database specific subject headings or index terms:

Search 5: Influenza Virus Vaccine Prediction **212 results**

Keywords, synonyms:
((((("influenza vaccines"[MeSH Terms] OR ("influenza"[All Fields] AND "vaccines"[All Fields])) OR "influenza vaccines"[All Fields]) OR (("influenza"[All Fields] AND "virus"[All Fields]) AND "vaccine"[All Fields])) OR "influenza virus vaccine"[All Fields]) AND ((((((((((((((("predict"[All Fields] OR "predictabilities"[All Fields]) OR "predictability"[All Fields]) OR "predictable"[All Fields]) OR "predictably"[All Fields]) OR "predicted"[All Fields]) OR "predicting"[All Fields]) OR "prediction"[All Fields]) OR "predictions"[All Fields]) OR "predictive"[All Fields]) OR "predictively"[All Fields]) OR "predictiveness"[All Fields]) OR "predictives"[All Fields]) OR "predictivities"[All Fields]) OR "predictivity"[All Fields]) OR "predicts"[All Fields])
"predict"[All Fields] OR "predictabilities"[All Fields] OR "predictability"[All Fields] OR "predictable"[All Fields] OR "predictably"[All Fields] OR "predicted"[All Fields] OR "predicting"[All Fields] OR "prediction"[All Fields] OR "predictions"[All Fields] OR "predictive"[All Fields] OR "predictively"[All Fields] OR "predictiveness"[All Fields] OR "predictives"[All Fields] OR "predictivities"[All Fields] OR "predictivity"[All Fields] OR "predicts"[All Fields]

Search 6: Influenza Virus Vaccine Database **286 results**

Keywords, synonyms:
((((("influenza vaccines"[MeSH Terms] OR ("influenza"[All Fields] AND "vaccines"[All Fields])) OR "influenza vaccines"[All Fields]) OR (("influenza"[All Fields] AND "virus"[All Fields]) AND "vaccine"[All Fields])) OR "influenza virus vaccine"[All Fields]) AND ((((("database"[All Fields] OR "database s"[All Fields]) OR "databased"[All Fields]) OR "databases"[All Fields]) OR "databasing"[All Fields])
Database specific subject headings or index terms:
"database"[All Fields] OR "database's"[All Fields] OR "databased"[All Fields] OR "databases"[All Fields] OR "databasing"[All Fields]

Search 7: Influenza Virus vaccine prediction (tools) and (models) **6 results**

Corpus:
((((((("influenza vaccines"[MeSH Terms] OR ("influenza"[All Fields] AND "vaccines"[All Fields])) OR "influenza vaccines"[All Fields]) OR (("influenza"[All Fields] AND "virus"[All Fields]) AND "vaccine"[All Fields])) OR "influenza virus vaccine"[All Fields]) AND ((((((((((((("predict"[All Fields] OR "predictabilities"[All Fields]) OR "predictability"[All Fields]) OR "predictable"[All Fields]) OR "predictably"[All Fields]) OR "predicted"[All Fields]) OR "predicting"[All Fields]) OR "prediction"[All Fields]) OR "predictions"[All Fields]) OR "predictive"[All Fields]) OR "predictively"[All Fields]) OR "predictiveness"[All Fields]) OR "predictives"[All Fields]) OR "predictivities"[All Fields]) OR "predictivity"[All Fields]) OR "predicts"[All Fields])) AND ("tool s"[All Fields] OR "tools"[All Fields])) AND ((((((((((((((((("model"[All Fields] OR "model s"[All Fields]) OR "modeled"[All Fields]) OR "modeler"[All Fields]) OR "modeler s"[All Fields]) OR "modelers"[All Fields]) OR "modeling"[All Fields]) OR "modelings"[All Fields]) OR "modelization"[All Fields]) OR "modelizations"[All Fields]) OR "modelize"[All Fields]) OR "modelized"[All Fields]) OR "modelled"[All Fields]) OR "modeller"[All Fields]) OR "modellers"[All Fields]) OR "modelling"[All Fields]) OR "modellings"[All Fields]) OR "models"[All Fields])

Studies Included:

1) **Models for Predicting the Evolution of Influenza to Inform Vaccine Strain Selection:**

   https://www.ncbi.nlm.nih.gov/pmc/articles/PMC5861780

- The topic for this review is partly similar to ours, but the overall research and approach is quite different. But we can refer to this as this is also doing vaccine strain selection.

Studies Excluded:

1) **Establishment of a Novel Safety Assessment Method for Vaccine Adjuvant Development**

   https://www.sciencedirect.com/science/article/pii/S0264410X18313616?via%3Dihub

   - Novel safety assessment method is not a part of our research.

2) **Model Comparisons for Effectiveness and the Cost Effectiveness of the vaccine:**
   https://www.sciencedirect.com/science/article/pii/S1098301518302985?via%3Dihub
   - We are not considering any cost related issues in our study.

3) **Modeling HIV Vaccine Trials of the future:**
   https://www.ncbi.nlm.nih.gov/pmc/articles/PMC5077275/
   - Our project is about influenza virus not HIV. Not Relevant.

4) **Using behavior change frameworks to improve healthcare worker influenza vaccination rates: A systematic review**
   https://www.sciencedirect.com/science/article/pii/S0264410X16302365?via%3Dihub
   - Not considered this as this is for only healthcare workers and we are studying the whole population in general.

5) **Biomarkers of safety and immune protection for genetically modified live attenuated leishmania vaccines against visceral leishmaniasis - discovery and implications.**
https://www.ncbi.nlm.nih.gov/pmc/articles/PMC4033241/
   - Irrelevant as our study is not based on leishmania.

Search 8: (influenza virus) OR (flu virus) OR (Viral Strain) AND (vaccine)) AND (prediction) AND (Database) **10 Results**

Word Corpus:
(((((((("orthomyxoviridae"[MeSH Terms] OR "orthomyxoviridae"[All Fields]) OR ("influenza"[All Fields] AND "virus"[All Fields])) OR "influenza virus"[All Fields]) OR ((((("influenza, human"[MeSH Terms] OR ("influenza"[All Fields] AND "human"[All Fields])) OR "human influenza"[All Fields]) OR "flu"[All Fields]) AND ((((((("virology"[MeSH Subheading] OR "virology"[All Fields]) OR "viruses"[All Fields]) OR "viruses"[MeSH Terms]) OR "virus s"[All Fields]) OR "viruse"[All Fields]) OR "virus"[All Fields])))) OR (((((("virally"[All Fields] OR "virals"[All Fields]) OR "virology"[MeSH Terms]) OR "virology"[All Fields]) OR "viral"[All Fields]) AND (((((("sprains and strains"[MeSH Terms] OR ("sprains"[All Fields] AND "strains"[All Fields])) OR "sprains and strains"[All Fields]) OR "strain"[All Fields]) OR "strains"[All Fields]) OR "strain s"[All Fields])))) AND ((((((((((((((((((((((((((((((((((("immune"[All Fields] OR "immuned"[All Fields]) OR "immunes"[All Fields]) OR "immunisation"[All Fields]) OR "vaccination"[MeSH Terms]) OR "vaccination"[All Fields]) OR "immunization"[All Fields]) OR "immunization"[MeSH Terms]) OR "immunisations"[All Fields]) OR "immunizations"[All Fields]) OR "immunise"[All Fields]) OR "immunised"[All Fields]) OR "immuniser"[All Fields]) OR "immunisers"[All Fields]) OR "immunising"[All Fields]) OR "immunities"[All Fields]) OR "immunity"[MeSH Terms]) OR "immunity"[All Fields]) OR "immunization s"[All Fields]) OR "immunize"[All Fields]) OR "immunized"[All Fields]) OR "immunizer"[All Fields]) OR "immunizers"[All Fields]) OR "immunizes"[All Fields]) OR "immunizing"[All Fields]) OR "vaccin"[Supplementary Concept]) OR "vaccin"[All Fields]) OR "vaccinable"[All Fields]) OR "vaccinal"[All Fields]) OR "vaccinate"[All Fields]) OR "vaccinated"[All Fields]) OR "vaccinates"[All Fields]) OR "vaccinating"[All Fields]) OR "vaccinations"[All Fields]) OR "vaccination s"[All Fields]) OR "vaccinator"[All Fields]) OR "vaccinators"[All Fields]) OR "vaccine s"[All Fields]) OR "vaccined"[All Fields]) OR "vaccines"[MeSH Terms]) OR "vaccines"[All Fields]) OR "vaccine"[All Fields]) OR "vaccins"[All Fields])) AND ((((((((((((((("predict"[All Fields] OR "predictabilities"[All Fields]) OR "predictability"[All Fields]) OR "predictable"[All Fields]) OR "predictably"[All Fields]) OR "predicted"[All Fields]) OR "predicting"[All Fields]) OR "prediction"[All Fields]) OR "predictions"[All Fields]) OR "predictive"[All Fields]) OR "predictively"[All Fields]) OR "predictiveness"[All Fields]) OR "predictives"[All Fields]) OR "predictivities"[All Fields]) OR "predictivity"[All Fields]) OR "predicts"[All Fields])) AND ((((("database"[All Fields] OR "database s"[All Fields]) OR "databased"[All Fields]) OR "databases"[All Fields]) OR "databasing"[All Fields])

Studies Included :

1) **Vaccines for preventing Influenza in healthy adults:**
   https://www.cochranelibrary.com/cdsr/doi/10.1002/14651858.CD001269.pub6/full
   - We will use the data of this clinical trial for making our prediction model.
   - It aims to identify, retrieve and assess all studies evaluating the effects (efficacy, effectiveness and harms) of vaccines against influenza in healthy adults.

2) **H7N3 live attenuated influenza vaccine has a potential to protect against new H7N9 avian influenza virus.**
   sciencedirect.com/science/article/pii/S0264410X13011316?via%3Dihub
   - This clinical might help us shape our prediction model in general, although it is for a particular case.
   -

3) **Bioinformatics Databases and Tools in Virology Research: An Overview**
   https://content.iospress.com/articles/in-silico-biology/isb00345
   (It provides a guide and overview of computational programs for analyses as a resource for genomics and proteomics studies in virology research. These resources are useful for understanding viral diseases, as well as for the design and development of antiviral agents.)

Studies Excluded:

1) **The determinants of 2009 pandemic A/H1N1 influenza vaccination: a systematic review.**
   https://www.sciencedirect.com/science/article/pii/S0264410X1102038X?via%3Dihub
   - Although our data is based on a large number of viruses, this analysis of a particular virus gives us a lot of insights about that virus.

2) **Prospects for engineering and improvement of cross-protective virus strains.**
   https://linkinghub.elsevier.com/retrieve/pii/S1879-6257(17)30058-5
   - Not relevant to flu virus vaccine strain prediction

3) **An overview of bioinformatics tools for epitope prediction: implications on vaccine development.**
https://www.sciencedirect.com/science/article/pii/S1532046414002330?via%3Dihub
   - This review is based on epitope prediction which is not in our scope as we have not considered any kind of pathway analysis.

4) **Interventions for autumn exacerbations of asthma in children**
https://www.ncbi.nlm.nih.gov/pmc/articles/PMC6494188/
   - Not relevant to our topic of interest.

5) **Cancer Immunoinformatics: A Promising Era in the Development of Peptide Vaccines for Human Papillomavirus-induced Cervical Cancer.**
http://www.eurekaselect.com/167061/article
   - Not relevant to our topic of interest(Cancer is not a part of our study).

6) **Diagnosis and management of febrile infants (0-3 months)**
https://www.ncbi.nlm.nih.gov/books/NBK92690/
   - Not relevant to flu virus vaccine strain prediction

7) **Applications for T-cell epitope queries and tools in the Immune Epitope Database and Analysis Resource.**
https://linkinghub.elsevier.com/retrieve/pii/S0022-1759(10)00322-4
   - This review is based on epitope prediction which is not in our scope as we have not considered any kind of pathway analysis.

Search 9: **6 results**

Influenza virus and vaccine and prediction tools and database and sequence

((("orthomyxoviridae"[MeSH Terms] OR "orthomyxoviridae"[All Fields]) OR ("influenza"[All Fields] AND "virus"[All Fields])) OR "influenza virus"[All Fields]) AND ((((((((((((((((((((((((((((((((("immune"[All Fields] OR "immuned"[All Fields]) OR "immunes"[All Fields]) OR "immunisation"[All Fields]) OR "vaccination"[MeSH Terms]) OR "vaccination"[All Fields]) OR "immunization"[All Fields]) OR "immunization"[MeSH Terms]) OR "immunisations"[All Fields]) OR "immunizations"[All Fields]) OR "immunise"[All Fields]) OR "immunised"[All Fields]) OR "immuniser"[All Fields]) OR

"immunisers"[All Fields]) OR "immunising"[All Fields]) OR "immunities"[All Fields]) OR "immunity"[MeSH Terms]) OR "immunity"[All Fields]) OR "immunization s"[All Fields]) OR "immunize"[All Fields]) OR "immunized"[All Fields]) OR "immunizer"[All Fields]) OR "immunizers"[All Fields]) OR "immunizes"[All Fields]) OR "immunizing"[All Fields]) OR "vaccin"[Supplementary Concept]) OR "vaccin"[All Fields]) OR "vaccinable"[All Fields]) OR "vaccinal"[All Fields]) OR "vaccinate"[All Fields]) OR "vaccinated"[All Fields]) OR "vaccinates"[All Fields]) OR "vaccinating"[All Fields]) OR "vaccinations"[All Fields]) OR "vaccination s"[All Fields]) OR "vaccinator"[All Fields]) OR "vaccinators"[All Fields]) OR "vaccine s"[All Fields]) OR "vaccined"[All Fields]) OR "vaccines"[MeSH Terms]) OR "vaccines"[All Fields]) OR "vaccine"[All Fields]) OR "vaccins"[All Fields]) AND ((((((((((((((("predict"[All Fields] OR "predictabilities"[All Fields]) OR "predictability"[All Fields]) OR "predictable"[All Fields]) OR "predictably"[All Fields]) OR "predicted"[All Fields]) OR "predicting"[All Fields]) OR "prediction"[All Fields]) OR "predictions"[All Fields]) OR "predictive"[All Fields]) OR "predictively"[All Fields]) OR "predictiveness"[All Fields]) OR "predictives"[All Fields]) OR "predictivities"[All Fields]) OR "predictivity"[All Fields]) OR "predicts"[All Fields]) AND ("tool s"[All Fields] OR "tools"[All Fields]) AND ((((("database"[All Fields] OR "database s"[All Fields]) OR "databased"[All Fields]) OR "databases"[All Fields]) OR "databasing"[All Fields]) AND ((((((((((((((("base sequence"[MeSH Terms] OR ("base"[All Fields] AND "sequence"[All Fields])) OR "base sequence"[All Fields]) OR "sequence"[All Fields]) OR "sequences"[All Fields]) OR "sequence analysis"[MeSH Terms]) OR ("sequence"[All Fields] AND "analysis"[All Fields])) OR "sequence analysis"[All Fields]) OR "sequencing"[All Fields]) OR "sequence s"[All Fields]) OR "sequenceable"[All Fields]) OR "sequenced"[All Fields]) OR "sequenceing"[All Fields]) OR "sequencer"[All Fields]) OR "sequencers"[All Fields]) OR "sequencies"[All Fields]) OR "sequencings"[All Fields])

Studies Included:

1) **Influenza Research Database: An integrated bioinformatics resource for influenza virus research**
   https://www.ncbi.nlm.nih.gov/pmc/articles/PMC5210613/
   (It mainly consists of influenza-related vaccines, sequence database which we will use in our project for prediction purposes)

2) **Bioinformatics Databases and Tools in Virology Research: An Overview**
   https://content.iospress.com/articles/in-silico-biology/isb00345
   - This study came as a repeat from the previous search.

Studies Excluded:

1) **Conservation Region Finding for Influenza A Viruses by Machine Learning Methods of N-linked Glycosylation Sites and B-cell Epitopes**
   https://www.sciencedirect.com/science/article/abs/pii/S0025556418305959?via%3Dihub
   - It has been excluded because this study is more closely related to pathway which is not included in our study and it is also related to vaccine invalidation which is different from our study.

2) **An Integrated Bioinformatics Approach to the Characterization of Influenza A/H5N1 Viral Sequences by Microarray Data: Implication for Monitoring H5N1 Emerging Strains and Designing Appropriate Influenza Vaccines**
   https://www.sciencedirect.com/science/article/abs/pii/S0890850810000630?via%3Dihub
   - It has been excluded because this study is more closely related to pathway and is also oriented towards only one type of influenza viruses i.e. H5N1 viral sequences but our study is related to prediction of new on basis of old database.

3) **Antibody-protein Interactions: Benchmark Datasets and Prediction Tools Evaluation**
   https://www.ncbi.nlm.nih.gov/pmc/articles/PMC2174481/
   - It has been excluded because our analysis is not considering 3D structures of protein and we are only considering the sequences and it is also using experimental data of X-ray crystallography which is not feasible for us to do due to certain constraints.

4) **Predicting Hemagglutinin MHC-II Ligand Analogues in Anti-TNFα Biologics: Implications for Immunogenicity of Pharmaceutical Proteins**
   https://www.ncbi.nlm.nih.gov/pmc/articles/PMC4536234/
   - It has been excluded because this study is more closely related to pathway and is also oriented towards development of anti-drug antibodies in patients with rheumatoid arthritis (RA) but our study is related to prediction of new on basis of old database.

Conclusion

In this project we are collecting viral strain data which can help in predicting a vaccine which can be used to treat if the victim or patient is suffering from this virus, to provide and make vaccines which can affect the population in advance,to find ways that can lead us to a universal vaccine.

Cited References

[1]https://www.cdc.gov/flu/about/burden/index.html
[2]https://www.ncbi.nlm.nih.gov/books/NBK22148/
[3]Clancy, S. (2008) Genetics of the influenza virus. Nature Education 1(1):83
[4]Susanna Esposito. (2018) 100 years since the 1918 influenza pandemic. Human Vaccines & Immunotherapeutics 14:3, pages 504-507.
[5]Carrat F, Flahault A. Influenza vaccine: the challenge of antigenic drift. Vaccine. 2007;25(39-40):6852–6862. doi:10.1016/j.vaccine.2007.07.027
[6]Nachbagauer R, Krammer F. Universal influenza virus vaccines and therapeutic antibodies. Clin Microbiol Infect. 2017;23(4):222–228. doi:10.1016/j.cmi.2017.02.009

Information Sources

Research Papers selected for project

1. Pubmed
   www.nature.com/articles/nrd4529
   Human Influenza Virus Infections
   Making Better Influenza Virus Vaccines?
   Influenza vaccine failure: Failure to protect or failure to understand?
   https://sci-hub.tw/10.1016/j.japh.2017.08.002
   Selecting Viruses for the Seasonal Influenza Vaccine
   Why do I need a flu vaccine every year?
   Immune History and Influenza Vaccine Effectiveness
   https://content.iospress.com/articles/in-silico-biology/isb00345
   https://www.ncbi.nlm.nih.gov/pmc/articles/PMC5861780

2. Google Scholar
   www.sciencedirect.com/science/article/pii/S0264410X07008328?via%3Dihub
   Current and future influenza vaccines
   10.1016/j.tim.2017.09.004
   https://www.cochranelibrary.com/cdsr/doi/10.1002/14651858.CD001269.pub6/full
   sciencedirect.com/science/article/pii/S0264410X13011316?via%3Dihub

Databases and relevant websites

https://www.fludb.org/brc/vaccineRecommend.spg?decorator=influenza
https://www.cdc.gov/flu/weekly/pastreports.htm
https://www.ncbi.nlm.nih.gov/gene

Databases

1. Genbank
2. FluDB (https://www.fludb.org/brc/home.spg?decorator=influenza)
3. CDC (https://www.cdc.gov/flu/weekly/pastreports.htm)

Web Search Engines

1) Google
2) Startpage