**“Why did I get the flu?”. Deep sequencing, error control, p-value, viral**

**evolution.**

Alekseeva L., Kureev N., Ovsyannikova A.

**Abstract**A vaccine is an effective way to get resistance to various diseases. It helps to get specific antibodies to protect organism from viral antigens. However, viruses can mutate quickly which give them ability to adapt to new environment. It causes many problems such as viral diseases of virial-resistant organisms. So our main aim is to figure out the reason of H2N3 influenza strain adaptation.

**Introduction**A vaccine is a preparation that provides immunity to diseases. It consists of wicked or killed forms of disease-causing organisms. These agents stimulate immune system to produce new antibodies that are able to interact with real viral particles [1]. The main mechanism that provides viral resistance is an antigenic drift. It allows viruses to adapt by using their genome replication mistakes, which couldn't be repaired because of lack viral reparation system and low quality of RNA-pol replication ability. This mechanism causes viral quasispecies to appear, when different polymorphic variants of the same organism interact and support each other [2].

**Materials and methods**Datasets we used were prepared by 2 ILLUMINA (Illumina MiSeq) runs (716,530 spots, 105.4M bases). Confirmed H3N2 influenza RNA from nasopharyngeal samples were extracted using Qiagen Viral RNA Mini Kits and underwent RT-PCR using custom primers for the HA gene. PCR products were purified using Qiagen PCR purification kits. Sequencing libraries were prepared from H3 amplicon with Nextera XT Sample Prep kit and Nextera XT 24 index kit. [3]

Raw reads were quality-checked with FastQC v. 0.11.8.We used BWA align program with BWA-MEM mode. The *influenza hemagglutinin gene* (KF848938.1) from NCBI database was used as a reference.With a help of Samtools and VarScan (with a minimum variant frequency of 0.001) we got positions where mutations were more likely to occur.Using aligned reads by BWA and annotation file from reference genome we got visual representation of SNP distribution by using IGV browser. Three more reads were aligned to reference data. The average and standard deviation of the frequencies from each reference sample was shown in Jupyter Notebook [4].

**Results**

There were in total 358265, 256586, 233327, 249964 reads for SRR1705851, SRR1705858, SRR1705859, SRR1705860, respectively.

A total of 21 SNPs were detected with a minimum frequency of 0.001. Compared to the *influenza hemagglutinin gene*, there were five SNPs, represented in the viral population (Table 1). Other were rare mutations.

*Table 1. The frequent SNPs analysis results*

|  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- |
| №SNP | SNP position | Base | Triplet | Amino Acid | Frequency |
| 1 | 72 | A —> G | CAA —> CGA | Val —> Ala | 99,96% |
| 2 | 117 | C —> T | CCA —> CTA | Gly —> Asp | 99,82% |
| 3 | 774 | T —> C | TTA —> TCA | Asn —> Ser | 99,97% |
| 4 | 999 | C —> T | GCG —> GTG | Arg —> His | 99,88% |
| 5 | 1260 | A —> C | TAT —> TCT | Ile —> Arg | 99,94% |

We identified one point mutation at residue 332 (SNP №4), which is related to Epitope C (residues 328–332, 334, 336, 338, 339, 341–344, 346, 347, 357–359, 366–370). Notably, there were all of the polymorphisms led to the nonsynonymous mutations(Table 1).

**Discussion**

Hemagglutinin is one of the proteins that allow the influenza virus to attach to the host cell [5]. Also, this protein is one of the main factors in the adaptation of the virus to the human population. Mutations occurring in the sequences responsible for the composition of the epitope allow the flu to create binding sites for new cells.

In our current results, we identified mutations in regions of HA that confer virus neutralization. Based on the data obtained, we assume that all of the mutations could affect the fact that the flu vaccine did not work, one of which is located in the epitope C [6].

Conducting a deeper analysis to obtain and compare found mutations to the mutations of three reference sequences, we were able to suggest their probable genesis. Number of SNPs with frequency more than 0.29 appeared as a result of amplification errors and other SNPs with a lower frequency seem to be sequencing errors. This ratio may be due to the fact that during the amplification process, the final product accumulates along with the mutations.

**References**

1. Melief CJ, van Hall T, Arens R, Ossendorp F, van der Burg SH. Therapeutic cancer vaccines //*J Clin Invest*. - 2015. - ;125(9):3401–3412. doi:10.1172/JCI80009

2. Carrat, F., & Flahault, A. (2007). *Influenza vaccine: The challenge of antigenic drift. Vaccine, 25(39-40), 6852–6862.* doi:10.1016/j.vaccine.2007.07.027

3. <https://www.ncbi.nlm.nih.gov/sra/?term=SRR1705851>

4. [Project log](https://github.com/alexmorphine/bioinformatics_workshops/blob/master/Project2/Project2-log.ipynb) (Jupyter Notebook)

5. Krammer, F., & Palese, P. (2013). *Influenza virus hemagglutinin stalk-based antibodies and vaccines. Current Opinion in Virology, 3(5), 521–530.* doi:10.1016/j.coviro.2013.07.007

6. Munoz E. T., Deem M. W. Epitope analysis for influenza vaccine design //Vaccine. – 2005. – Т. 23. – №. 9. – С. 1144-1148.