Probable causes of flu development in a vaccinated person

Abstract

Single nucleotide variations inside separate populations of viruses play a crucial role in virus evolution and vaccine resistance development, whereas it is barely detectable in early stages when using hemagglutination assays or conventional sequencing techniques. This project is dedicated to application of deep sequencing and bioinformatics analysis for virus antigenic drift discovery. Here we investigated the deep sequencing results of Influenza A virus strain that infected patient vaccinated against this pathogen in the 2014 season. 7 statistically significant SNPs were observed whereas only one comparably rare (0,83%) missense mutation Pro103Ser was capable of changing in Epitope D providing evidence for effectiveness of this technology.

Introduction

Antigenic drift is a type of genetic variability in viruses that occurs due to the accumulation of mutations in the virus genes encoding surface proteins recognized by host antibodies. One of the antigens subject to antigenic drift in the influenza virus is hemagglutinin. This is a surface glycoprotein that allows the virus to bind to the sialic acid of the cell and to attach to the host cell membrane. The regions of hemagglutinin recognized by the host immune system are under constant selective pressure due to the high prevalence of influenza and a certain level of vaccination of the population. That leads to imminent glycoprotein modifications inside the sick host where these mutations are not necessary and not so abundant while subsequent transmission to the vaccinated people can possess virus to find a new ecologic niche and spread [1]. Intensive development of sequencing techniques opens new horizons for today's science and medicine. Oldfashioned techniques such as hemagglutination and classical sequencing can't recognize such sparse virus emergence in unvaccinated patients while deep sequencing does. Due to high resolution this method allows us to distinguish between rarely presented "real" SNPs and sequencing errors in order to estimate significance of these mutations and predict possible outcomes.

Methods

The Influenza A virus (A/USA/RVD1_H3/2011(H3N2)) segment 4 hemagglutinin (HA) gene from NCBI was used as a reference genome [2]. Raw Illumina sequencing reads from shotgun sequencing of HA of the virus that infected the neighbor (my sick and contagious neighbor, ms&cn) were taken from SRA (SRR1705851) [2].

The quality of the readings was assessed using Trimmomatic (v. 0.39) [3].

Aligning sequence to reference was done using aligner BWA-MEM [4]. Samtools (v. 1.21) were used to work with alignment data and to make a mpileup file (with --max-depth 35000) [5].

Varscan (v. 2.4.0) was used to find variant positions with minimum variant allele frequency threshold=0.95 for frequently occurring variants and threshold=0.001 for rare [6]. IGV_2.18.4 was used for visualization [7].

Results

Since in order to observe the evolution of the virus in one organism it is necessary to find not only mutations that are common in the sample, but also rare ones, it is necessary to filter out real rare mutations from sequencing and amplification errors. To do this, triple deep sequencing data from an isogenic sample of the influenza virus H3N2 (C58, C59, C60) were analyzed (Table 1).

Table1. Triple deep sequencing

	C58	C59	C60
The number of started reads, b.p.	256586	233327	249964
The number of mapped reads, b.p.	256500	233251	249888
Average mutation's frequency, %	0,26	0,24	0,26
Standard deviation of the frequencies, %	0,07	0,05	0,08

Based on the fact that the control sample data contain only mutations associated with sequencing errors, a frequency threshold was determined above which a mutation can be considered non-random: average of average mutation's frequency + average of standard deviation of the frequencies * 3= 0,25% + 0,07%*3 = 0,46%.

The data ms&cn is presented in 358265 readings (358032 mapped) of acceptable quality. Analysis showed the presence of 7 single nucleotide polymorphisms with frequency > 0,46% in the sample ms&cn (Table 2)

Table 2. Single nucleotide polymorphisms found in ms&cn

Position in genome	Original -> Altered nucleotide	Frequency	Mutation variant
72	A -> G	99,96	synonymous
117	C -> T	99,82	synonymous
307	C -> T	0,95	synonymous
774	T -> C	99,97	synonymous
999	C -> T	99,86	synonymous
1260	A -> C	99,94	synonymous
1458	T -> C	0,83	missense Pro103Ser

Nucleotide change on positions 307 did not affect the codon, because “TAT” which codes Tyrosine was replaced with the other Tyrosine encoded by TAC as Lysine at position 1260, PhenylAlanine at 774, Proline at 117 and Threonine at 72.

Cytosine replaced by Thymine at position 999 possess an exchange of Alanine and Valine, which are both non-polar.

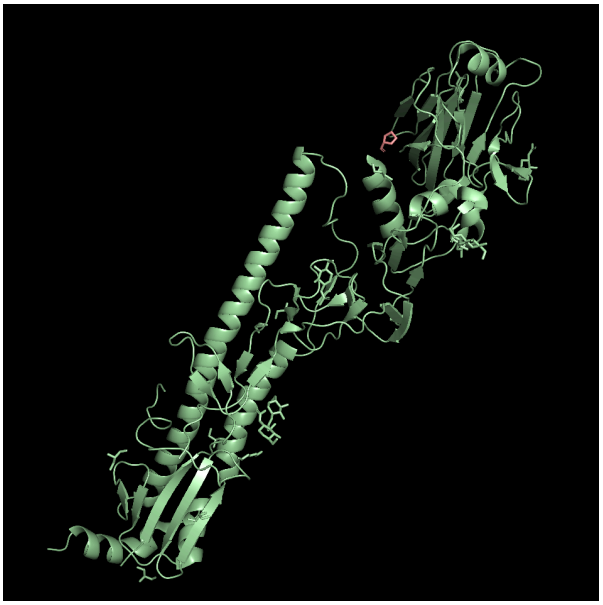

Figure 1. Crystal structure of A/Victoria/361/2011 (H3N2) influenza virus hemagglutinin (PDB 4O5N), 103 amino acid residue is highlighted in pink

Discussion

During the work, one missense mutation (Pro103Ser) was discovered, which occurs with a frequency of 0.83% in the studied sample. This amino acid residue is part of the epitope D of the influenza hemagglutinin [8].

According to a codon table, other SNPs appear to be synonymous as they do not affect the amino acid sequence or its characteristics, keeping the same polarity and acid-base properties despite the higher frequencies. Thus these mutations are not likely to be the reason for the patient’s flu.

There are other reasons that could lead to the disease. Individual immune responses vary, and some may not generate sufficient antibodies despite vaccination.

Working with such a sensitive type of experiments it’s crucial to minimize error probability. At least it’s possible to lower the number of PCR cycles and use high-fidelity DNA polymerases with low error rates or use combined sequencing techniques utilizing both short and long reads.

As bioinformatics we are able to implement tools such as LoFreq or GATK's Base Quality Score Recalibration (BQSR), which use sequence quality metrics and models to correct systematic errors. Also we can use package qvalue for control of False Positive results, capable of multiple comparison amendments and put less strict borders for our variant caller.

References

1. Carrat F., Flahault A. Influenza vaccine: the challenge of antigenic drift //Vaccine. – 2007. – Т. 25. – №. 39-40. – С. 6852-6862. https://doi.org/10.1016/j.vaccine.2007.07.027
2. Sayers, E. W., Bolton, E. E., Brister, J. R., Canese, K., Chan, J., Comeau, D. C., Connor, R., Funk, K., Kelly, C., Kim, S., Madej, T., Marchler-Bauer, A., Lanczycki, C., Lathrop, S., Lu, Z., Thibaud-Nissen, F., Murphy, T., Phan, L., Skripchenko, Y., … Sherry, S. T. (2022). Database resources of the national center for biotechnology information. *Nucleic Acids Research*, 50(D1), D20–D26. https://doi.org/10.1093/NAR/GKAB1112
3. Bolger, A. M., Lohse, M., & Usadel, B. (2014). Trimmomatic: a flexible trimmer for Illumina sequence data. *Bioinformatics*, 30(15), 2114. https://doi.org/10.1093/BIOINFORMATICS/BTU170
4. Li, H. (2013). *Aligning sequence reads, clone sequences and assembly contigs with BWA-MEM*. https://arxiv.org/abs/1303.3997v2
5. Danecek, P., Bonfield, J. K., Liddle, J., Marshall, J., Ohan, V., Pollard, M. O., Whitwham, A., Keane, T., McCarthy, S. A., & Davies, R. M. (2021). Twelve years of SAMtools and BCFtools. *GigaScience*, 10(2). https://doi.org/10.1093/GIGASCIENCE/GIAB008
6. Koboldt, D. C., Chen, K., Wylie, T., Larson, D. E., McLellan, M. D., Mardis, E. R., Weinstock, G. M., Wilson, R. K., & Ding, L. (2009). VarScan: variant detection in massively parallel sequencing of individual and pooled samples. *Bioinformatics*, 25(17), 2283. https://doi.org/10.1093/BIOINFORMATICS/BTP373
7. Thorvaldsdóttir H., Robinson J. T., Mesirov J. P. Integrative Genomics Viewer (IGV): high-performance genomics data visualization and exploration //Briefings in bioinformatics. – 2013. – Т. 14. – №. 2. – С. 178-192.
8. Muñoz, E. T., & Deem, M. W. (2005). Epitope analysis for influenza vaccine design. *Vaccine*, 23(9), 1144–1148. https://doi.org/10.1016/J.VACCINE.2004.08.028