Detection of the rare variants for analyzing hemagglutinin genes

Ignat Sonets, Kamilla Faizullina

We analyze the amplicon data to detect rare variants. As occuring errors exist, we use three control sequences to understand which variants are real and which of them was detected as mutation due to error.

1 Introduction

Vaccine is a formulation of killed or attenuated pathogens, or antigens derived from them which help prevent, ameliorate, or treat infectious disease by stimulating antibody production or cellular immunity against the pathogen [1]. Epitope is a molecular region, usually an amino acid sequence, on the surface of an antigen that is capable of eliciting a specific immune response. [2]. Any changes of epitopes lead to decreasing efficiency of antibodies.

Influenza viruses are important human respiratory tractpathogens responsible for the seasonal epidemics andsporadic pandemics around the world [4] Influenza exists as quasispecies. Quasispecies are viral variants leading to diversification of the original strain [3]. Deep sequencing enables us to study mixed populations. However, the detection of rare varints might be challenging due to errors which accure prior to or during sequencing.

2 Methods

2.1 Data

In order to analyze the hemaaglutinin genes, we use Amplicon of H3N2 HA infecting Homosapien labeled SRR1705851 from The National Center for Biotechnology Information database [5]. We analyze the quality of the amplicon data via Fastqc [6]. Figure 1 illustrates the quality of this sequencing data. It is quite good and we do not need to filter the reads. We use the reference data [9]. Also, we use three control sequencing data labeled SRR1705858, SRR1705859 and SRR1705860 [10].

The sequence	Reads	Mapped
SRR1705851	358265	
SRR1705858	256586	
SRR1705859	233327	
SRR1705860	249964	

2.2 Methods

In order to to map the data from the resistant strain to the reference sequence, we use the aligner called BWA-MEM [7]. VarScan is used to find SNP [8].

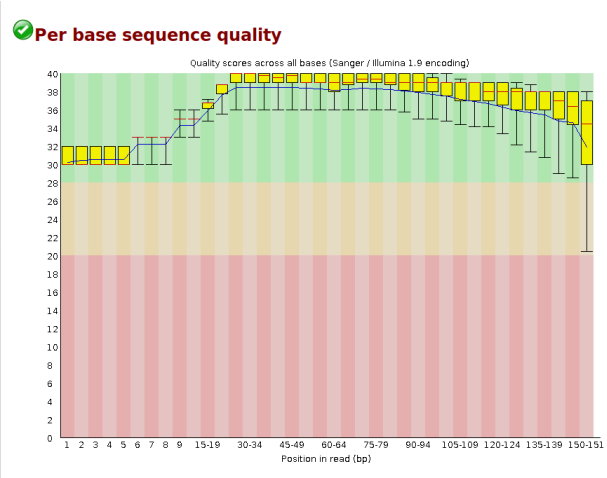

Figure 1: The quality of the sequence labeled SRR1705851

3 Results

3.1 Common variants

The data labeled SRR1705851 is used to detect the variants. The parameter –min-var-frequency equals 0.95 for searching common variants. Five SNPs were reported (Table 1).

3.2 Rare variants

So, We run VarScan again, but –min-var-frequency equals 0.001.

4 Discussion

Position	Ref	Alt	Triplets	Altered aminoacid
72	A	G	CAA → CGA	
117	C	T	CCA → CTA	fenilalanina/fenilalanina
774	T	C	TTA → TCA	Glutamine/Glutamine
999	C	T	GCG → GTG	leicins/leicins
1260	A	C	TAT → TCT	Threonine/Threonine

Table 1: The variants, –min-var-frequency=0.95

Position	Ref	Alt	Altered aminoacid	Frequency
307	C	T	Proline/Serine	0.9

Table 2: The variants, –min-var-frequency=0.001

References

[1] Vaccine. In: Encyclopedic Reference of Genomics and Proteomics in Molecular Medicine. Springer, Berlin, Heidelberg, 2006.

[2] Epitope. In: Vohr HW. (eds) Encyclopedia of Immunotoxicology. Springer, Berlin, Heidelberg, 2016.

[3] Quasispecies. In: Schwab M. (eds) Encyclopedia of Cancer. Springer, Berlin, Heidelberg, 2008.

[4] H. Ghaffari, A. Tavakoli, A. Moradi, A. Tabarraei, F. Bokharaei-Salim, M. Zahmatkeshan, M. Farahmand, D. Javanmard, S. J. Kiani, M. Esghaei, V. Pirhajati-Mahabadi, S. H. Monavari, and A. Ataei-Pirkooh, "Inhibition of h1n1 influenza virus infection by zinc oxide nanoparticles: another emerging application of nanomedicine," Journal of Biomedical Science , vol. 26, p. 70, Sep 2019

[5] NCBI: https://www.ncbi.nlm.nih.gov/sra/?term=SRR1705851

[6] Fastqc : https://www.bioinformatics.babraham.ac.uk/projects/fastqc/

[7] Burrows-Wheeler Aligner: http://bio-bwa.sourceforge.net/

[8] VarScan: http://dkoboldt.github.io/varscan/

[9] Reference data: http://public.dobzhanskycenter.ru/mrayko/Week2/KF848938.1.fasta

[10] SRR1705858: https://trace.ncbi.nlm.nih.gov/Traces/sra/?run=SRR1705858
SRR1705859: https://trace.ncbi.nlm.nih.gov/Traces/sra/?run=SRR1705859
SRR1705860: https://trace.ncbi.nlm.nih.gov/Traces/sra/?run=SRR1705860

[11] Munoz E. T., Deem M. W. Epitope analysis for influenza vaccine design,Vaccine,2005.