BIOINFORMATICS INSTITUTE

Rare variant identification with the help of deep sequencing.

December 7, 2023

Homework number 3

by Sapozhnikov N.A.

1 Abstract

A scenario where despite receiving the flu vaccine, an individual falls ill with symptoms matching the flu after contacting with unvaccinated person contracts the virus is a quite rare one. To investigate the discrepancy, a hemagglutination inhibition (HI) assay is conducted, revealing a close match between the roommate's virus and the H3N2 strain A/Hong Kong/4801/2014. Suspecting viral quasispecies, the individual opts for a targeted deep sequencing experiment using Illumina single-end sequencing to analyze the hemagglutinin (HA) genes in the viral sample.

2 Introduction

In the dynamic landscape of influenza prevention, breakthrough infections challenge the effectiveness of vaccination strategies. This narrative unfolds with a vaccinated individual experiencing flu-like symptoms post-exposure to an unvaccinated roommate's virus. Utilizing a hemagglutination inhibition (HI) assay, we identify a close match between the roommate's viral strain and H3N2 variant A/Hong Kong/4801/2014, prompting questions about vaccine coverage. To delve deeper into viral dynamics, a targeted deep sequencing analysis of hemagglutinin (HA) genes is initiated using Illumina single-end sequencing. This investigation aims to elucidate the potential role of viral quasispecies in breakthrough infections. Our study contributes to the ongoing discourse on influenza prevention, emphasizing the interplay between vaccination strategies and viral evolution within hosts.

3 Methods

A comprehensive workflow, implemented in the Snakemake framework, was employed to facilitate the analysis of deep sequencing data in this study. The workflow encompassed a series of tools designed for handling high-throughput sequencing information, including BWA (Burrows-Wheeler Aligner) and VarScan. The workflow seamlessly integrated BWA and VarScan into a unified analytical framework, streamlining the entire process from read alignment to variant calling. The modularity of the pipeline facilitated adaptability for future analyses and extensions. By employing this Snakemake-based workflow, coupled with tools such as BWA and VarScan, we aimed to conduct a thorough and systematic analysis of deep sequencing data, unraveling the genomic dynamics underlying the breakthrough influenza infection.

4 Results

Two variants 1458 T -> C and 307 C -> T were found with frequencies 0.94 and 0.84. Among all mutations, the one on position 307 was non-synonymous. This mutation affects the epitope region D in the hemagglutinin protein as has been reported in another work [1].

5 Discussion

To further enhance the accuracy of identifying and quantifying rare variants in deep sequencing experiments, one additional approach to control for error involves the incorporation of Unique Molecular Identifiers (UMIs). UMIs are short sequences that are ligated to individual DNA or RNA molecules before the PCR amplification step. By assigning a unique identifier to each original template molecule, UMIs enable the distinction between PCR duplicates and true biological variants. This approach helps mitigate errors introduced during PCR amplification, a critical step in the sequencing process [2].

The implementation of UMIs provides several advantages in error control. Firstly, it allows for the identification and removal of PCR duplicates, which are a common source of artifacts in deep sequencing data. Additionally, UMIs assist in the accurate quantification of rare variants by reducing the impact of PCR bias, ensuring a more faithful representation of the true viral population. By incorporating UMIs into our experimental workflow, we can significantly enhance the precision and reliability of variant calling, particularly for low-frequency variants.

6 Bibliography

References

[1] E. T. Muñoz and M. W. Deem, "Epitope analysis for influenza vaccine design," *Vaccine*, vol. 23, pp. 1144–1148, Jan. 2005.

[2] T. Smith, A. Heger, and I. Sudbery, "UMI-tools: modeling sequencing errors in Unique Molecular Identifiers to improve quantification accuracy," *Genome Research*, vol. 27, pp. 491–499, Mar. 2017.