DOI 10.18699/SBB-2022-517

The potential usability of Oxford Nanopore Technology for revealing of SARS-CoV-2 dual infection and intra-host viral variability in clinical samples

Penkin L.N.*, Korneenko E.V., Samoilov A.E., Manolov A.I., Speranskaya A.S.
Research Institute for Systems Biology and Medicine of the Federal Service for Surveillance on Consumer Rights Protection and Human Wellbeing, Moscow, Russia
** leopold.valerjanovitch@yandex.ru*

Key words: SARS-CoV-2, Oxford Nanopore, heterogeneity, intra-host variability, double infection

Motivation and Aim: Due to the emergence of the epidemic of COVID-19 caused by SARS-CoV-2 it has become necessary to monitor the spread of different variants of the virus. Since its appearance, the virus has continued to evolve, giving rise to many variants with increased transmissibility and ability to avoid the immune system. The patient's organism presents the reservoir for the new strains due to ongoing viral evolution. In addition, the facts of Covid-19 double infection(the case when two strains are simultaneously detected in the patient [1]) and their impact on the course of the disease has not been sufficiently studied) has been described repeatedly. Therefore, it is important not only to be able to identify the strains of SARS-CoV-2, but also to identify the variability in samples and determine its nature as well. The most informative method of determining SARS-CoV-2 variants is high-throughput sequencing. Currently, a large amount of data of the SARS-CoV-2 virus genome has been obtained using Oxford Nanopore Technology (ONT). But, according to the public opinion, this technology produces a lot of mistakes. Therefore, this fact raises the question, if it's possible to distinguish the experimental artifacts from the true intra-host viral variability. True heterogeneity refers to positions that result from the presence of several virus variants in the sample, but are not the result of contamination or errors produced by polymerases, reverse transcriptases or sequencing errors.
The aim of our study is to estimate the usability of Oxford Nanopore data for detection of intra-host variability and double infection of SARS-CoV-2.
Methods and Algorithms: Samples for sequencing (oropharyngeal swabs) were obtained from 32 regions of the Russian Federation from patients positive for SARS-Cov-2 and collected from January 19, 2022 to March 30, 2022. RNA was extracted by *NA magnet SARS-CoV-2 (LYTECH)*, the whole genome sequencing was performed by the Oxford Nanopore platform (PromethIon, GridIon) using Midnight RT PCR Expansion Kit (EXP-MRT001, ONT) and Rapid Barcoding kit 96 (SQK-RBK-110-96, ONT). Consensus sequences were obtained using epi2me-labs/wf-artic utilities. Detection of the virus lineage was performed using the Pangolin algorithm. Variable positions were detected using Longshot software - is a variant calling tool for diploid genomes using long error prone reads such as Pacific Biosciences (PacBio) SMRT and Oxford Nanopore Technologies (ONT).
Results: The whole genome sequencing and detection of viral strain were performed for 6916 samples. The distribution of variants was as follows: Omicron subvariant BA.1 – 5310, Omicron subvariant BA.2 – 1028, Delta – 280, None – 257. The search for SNV in the samples was performed in several stages. At first, we used the epi2me-labs/wf-

artic algorithm with default settings (sample coverage normalization threshold = 200). As a result, the usage of the default settings revealed an artificial distribution of the frequency of occurrence of SNV in the samples (anomalous "clustering"). By increasing the maximum coverage threshold to 5000, this artifact was eliminated. Then, we identified samples in which reliable SNVs were present. The positions in which the minor variant was represented by at least in 35–50 % of the reads were taken as reliably defined SNVs. As a result, in 3,366 samples at least one valid SNV was detected. At the same time, 95 % of these samples contained 1 to 10 SNV (N = 3183). In the remaining 5 % (N = 35) of samples, between 11 and 153 SNV were found and these samples were further analyzed to determine if the nature and position of the SNV could be explained by the fact that these samples could potentially contain two different strains of the virus. The samples in our study relate to the period with the following distribution of SARS-Cov-2 variants: B.1.1.617.2 (Delta), BA.1 (Omicron), BA.2 (stealth-Omicron). Therefore, a list of mutations in the S-gene for Delta/BA.1, Delta/BA.2, and BA.1/BA.2 variants was constructed. In 4 samples, the degeneracy and SNV position corresponded to a mixture of Delta and Omicron strains. Additional analysis suggests that this variability is not the result of intra-laboratory contamination. Probably, the patients from whom these samples were obtained were simultaneously infected with two strains of SARS-CoV-2.

Conclusion: Our analyses revealed that 48.6 % of samples contained at least one SNV with low frequency variants in 35–50 % of allele frequency. We suppose the Oxford Nanopore sequencing data could be used for investigation of intra-host variation and double infection of SARS-CoV-2 at least for samples with high viral load of both strains.

Acknowledgements: This work was supported by Rospotrebnadzor service state task No. 122030900051-9.

References

1. Samoilov A.E. et al. Case report: change of dominant strain during dual SARS-CoV-2 infection. *BMC Infect Dis*. 2021;21(1):959.