# Bioinformatics workflow for reliable detection of SARS-CoV-2 variants in wastewater data generated by massively parallel sequencing

M. Böhmer1,2, D. Čárska3, R. Hekel1,4,5, T. Sládeček1, J. Sitarčík1,3,5, A. Goga1,3, W. Krampl1,4, D. Rusňáková1,2,4, T. Sedláčková1, A. Kaliňáková2, J. Budiš1,5, T. Szemes1,2,4,5

*1 Comenius University Science Park, Bratislava, Slovakia*

*2 Public Health Authority of the Slovak Republic, Bratislava, Slovakia*

*3 Department of Computer Science, Faculty of Mathematics, Physics and Informatics, Comenius University, Bratislava, Slovakia*

*4 Department of Molecular Biology, Faculty of Natural Sciences, Comenius University,*

*Bratislava, Slovakia*

*5 Slovak Centre of Scientific and Technical Information, Bratislava, Slovakia*

The ongoing SARS-CoV-2 pandemic has caused the deaths of more than 6.2 million people around the world according to the World Health Organisation. It has urged us to find a way to predict emerging hotspots and therefore, with good precautionary measures, reduce the possibility of new cases and following deaths. Monitoring of wastewaters with the inclusion of massively parallel sequencing of wastewater samples appears to be an excellent tool for this purpose. In fact, novel SARS-CoV-2 variants from wastewater could be identified 1 to 2 weeks before being detected in clinical samples from the same area. The problem arises with analysing such data. In this work, we present a highly scalable and easily deployable computational pipeline for reliable detection of SARS-CoV-2 variants of concern in wastewater data generated by massively parallel sequencing. It uses state-of-the-art tools freely available online combined with our developed method to distinguish between SARS-CoV-2 variants of concern. Our bioinformatics workflow is currently used to monitor wastewater for SARS-CoV-2 variants in the Slovak republic which is coordinated by the national Public health authority.