An artificial intelligence tool to evaluate properties of SARS-CoV-2 mutations

Do Young Seong, Wonjong Choi, Seungpil Jeong, Jongkeun Park, Dongwan Hong

(ds528491@catholic.ac.kr)

Department of Medical Informatics, Graduate School, The Catholic University of Korea, Seoul, Korea

Artificial Intelligence analytics toolkit for predicting Virus mutation in protEin (AIVE; https://ai-ve.org) is a web and GPU based platform that hosts protein structure prediction and properties of viral sequences. AIVE hosts Amino acid Property Eigen Selection Score (APESS), a tool that is based on biochemical properties and protein structure. APESS is based on four preceding calculations: sub-clustering of protein structure (SCPS), polarity score (PS), mutation rate (MR), and biochemical properties eigen score (BPES). APESS evaluates the risk of SARS-CoV-2 lineages from known sequences, randomly sampled SARS-CoV-2 sequences, or sequences directly input by the user. Based on mutation data in SARS-CoV-2 variants, Alphafold2 is utilized to perform protein structure predictions. Using the Wuhan-hu-1 sequence as the reference, we performed comparative analysis for main VOCs: Alpha, Beta, Delta, and Omicron. Also, through visualization tools that provide Alphafold2 outputs such as pLDDT and PAE, reports are generated and displayed in AIVE. Monomer folding and multimer docking are also visualized and scored. Differences in pattern caused by mutations are shown for the input amino acid sequences based on polarity features and protein properties. APESS can predict the risk level of input sequences through the distribution of the values obtained through the computational modeling and using previously known lineages. We can visualize the results of the user’s sequence APESS score.