Dissecting the Dynamics of Human respiratory syncytial virus (hRSV) Protein Sequence Diversity

SUMMARY

Human respiratory syncytial virus (hRSV) is one of the most contagious viruses worldwide, which can lead to lower respiratory tract infections, such as bronchiolitis and pneumonia. There is no licensed prophylactic RSV vaccine. A large number of hRSV sequences are available in public databases, which were used in a large-scale proteome-wide sequence diversity analysis herein of hRSV and that led to the identification of potential vaccine targets for epitope-based hRSV vaccine design.

This study focussed on all available protein sequences of the virus. Sequence data was collected from publicly available NCBI Virus and Virus Pathogen Database and Analysis Resource (VIPR) databases. Protein sequence data were pooled, and duplicates were removed by use of CD-HIT. BLASTp search was performed on the collected sequences using UniProt reference records of the 11 hRSV proteins to generate individual protein datasets. Each protein dataset was multiple sequence aligned using MUSCLE. Sequence diversity was measured by use of Shannon's entropy for each overlapping 9-mer (nonamer) (1–9, 2–10, etc.) position across the length of the protein. From there on, T-cell epitopes were matched to experimentally validated ones in the IEDB database and were also predicted using the tools recommended by the IEDB resource benchmark. To support our prediction results, molecular docking experiments of selected epitopes against human leucocyte antigen (HLA) I and II and molecular dynamic simulations were performed.

The total number of sequences collected from VIPR and NCBI was 93,382. After the removal of duplicates, the number decreased to 12,413 (an ~ 87% reduction). The peak entropy value observed for hRSV was ~4.2 in the M2-2 protein at starting nonamer position 43. However, the proteome-wide average entropy was low (~0.8) and thus, indicating high conservation, with values ranging from ~0.5 (protein N) to ~2.3 (protein M2-2). Most of the variants (~55%) were between the entropy range of 0 to 1. Approximately half (~51%) of the nonamer positions were classified as highly conserved (index incidence ≥ 90%). The number of 2269 highly conserved nonamer sequences were selected for epitope prediction. From the 2269 nonamer sequences, 997 and 235 epitopes were predicted to be restricted to HLA Class I and II, respectively.

This study provides information about sequence diversity across the proteome of hRSV and identified potential vaccine targets.. The high conservation of the proteome and their predicted epitopes merits further investigation for prophylactic hRSV vaccine design.

Keywords: Viral sequence analysis, immunoinformatics, vaccine target discovery