**Supplementary Materials**

# Diversity and evolution of computationally predicted T cell epitopes against Human Respiratory Syncytial virus

Jiani Chen [1][2][3][4], Swan Tan [1][3][4][5], Vasanthi Avadhanula [6], Leonard Moise [3][7], Pedro A Piedra [6], Anne S De Groot [3][7], Justin Bahl [1][2][3][4][5][8] \*

\* Corresponding author email: [justin.bahl@uga.edu](mailto:justin.bahl@uga.edu)

**This file includes:**

Supplementary Figure 1 to Supplementary Figure 7

**Other Supplementary Materials for this manuscript include the following:**

Supplementary Data File 1

Timeline

Description automatically generated with medium confidence **Supplementary Figure 1. Distribution and diversity of T cell epitopes in RSV F protein.** The tree panel on the left is a time-scaled phylogeny build with RSV-A **(A)** or RSV-B **(B)** F gene nucleotide sequences using the ML approach. Determined genotypes are labeled on the right with black bars. Each color column on the right side represents the presence of an MHC class I or class II epitope. Only the epitopes that are present in more than 1% of sampled isolates are displayed. The column color indicates different numbers of epitope sequences at the same location.

A picture containing schematic

Description automatically generated **Supplementary Figure 2. Distribution and diversity of T cell epitopes in RSV G protein.** The tree panel on the left is a time-scaled phylogeny build with RSV-A **(A)** or RSV-B **(B)** G gene nucleotide sequences using the ML approach. The clades that contain novel 72-nt or 60-nt duplication at the second hypervariable region of G gene were highlighted in red. Determined genotypes are labeled on the right with black bars. Each color column on the right side represents the presence of an MHC class I or class II epitope. Only the epitopes that are present in more than 1% of sampled isolates were displayed. The column color indicates different numbers of epitope sequences at the same location.

Graphical user interface

Description automatically generated

**Supplementary Figure 3. Distribution of JanusMatrix Human Homology score for putative RSV MHC class I and class II epitopes.** The cross-reactive potential of identified putative T cell epitopes and human host was represented with a JanusMatrix Human Homology score. 6.45% identified putative class I epitopes and 1.12% class II epitopes are cross-conserved on the TCR face with human epitopes.

Calendar, scatter chart

Description automatically generated

Supplementary Figure 4. Predicted T cell epitope landscapes of RSV surface proteins. RSV T cell epitope landscapes were built with sequenced-based MHC class I epitope binding prediction (left), MHC class II epitope binding prediction (middle) or combining class I and class II epitope biding prediction (right). Sequences are colored by the epitope cluster determined by epitope landscapes built with combining Class I and Class II epitope prediction



**Supplementary Figure 5. Total within sum of squares (wss) using *k-means* algorithm.** Totals within sum of squares in epitope topographies were calculated after clustering into k (from 1 to 10) groups with *k-means*. The optimal number of clusters is determined to be 3 in the analysis of RSV-A F and G proteins and is determined to be 2 in the analysis of RSV-B F and G proteins using the Elbow method.

Graphical user interface, diagram

Description automatically generated

**Supplementary Figure 6. Validation of T cell epitope distance estimation using the IEDB analysis resource**. Validation is performed with MHC class I epitope binding prediction of RSV-A F protein. **(A)** Heatmaps for pairwise MHC class I epitope distance estimated in iVAX toolkits or calculated with custom python scripts using MHC class I molecule binding prediction that is implemented in IEDB. **(B)** Eigenvalues for each sequence are calculated from pairwise distance matrices using “RSpectra” package in R. The Pearson correlation test significantly supports a non-zero correlation between T cell epitope distance estimated with EpiCC and T cell epitope distance estimated with IEDB. **(C)** T cell epitope topographies are built with pairwise epitope distances estimated from EpiCC or IEDB. Both methods resulted in a similar cluster pattern for the CD8 T cell epitope profile of RSV-A F protein.



**Supplementary Figure 7. Evaluation of RSV vaccine candidate strains with T cell epitope content in different WHO regions**. RSV-A and RSV-B major surface protein sequences were grouped by isolation year and 6 isolated WHO regions, African Region (AFRO), Region of the Americas (PAHO), South-East Asia Region (SEARO), European Region (EURO), Eastern Mediterranean Region (EMRO) and Western Pacific Region (WPRO). Each year group was labeled by the latest isolated year of sequences after the previous group label. The proportion of cross-conserved T cell epitope content between vaccine strains (CP248 or CP52) and wild circulating strains in different year groups was represented by bar graphs.