**Input files section**

Epitope Analyzer is a Shiny app aimed to filter and analyze predicted T-cell epitopes. Epitope Analyzer includes 5 tools: (1) Epitope Distribution, (2) Epitope Density, (3) Epitope Location, (4) Epitope Promiscuity, and (5) Epitope Intersection. It needs a txt format file tab-delimited where the first 4 columns are the peptide sequence, the position of the peptide within the protein, the length of the protein, and the name of the protein. The following columns should be named with the allele containing the %rank score assigned to each peptide for each MHC allele. The “%rank score” is a number between 0 and 100. It is defined as the rank of the predicted binding score compared to a set of random natural peptides by several predictors of T-cell epitopes. By default, Epitope Analyzer considers 2 and 10 as the %rank cutoff for MHC Class I and MHC Class II epitopes, respectively. Usually, the output of these predictors requires to be parsed before uploading in the shiny app.

Once the input file is uploaded, parameters need to be set by the user followed by clicking on the “Run analysis” or “Plot” buttons. Each of the tools shows (1) graphics that can be downloaded as png files by clicking on the camera icon (“Download plot as png”), and (2) tables that can be downloaded by clicking on the “Download table” button.

We have included 3 example files that can be downloaded and used as input files on Epitope Analyzer.

* **File 1:** Predicted MHC Class I epitopes from SARS-CoV-2.
* **File 2:** Predicted MHC Class II epitopes from SARS-CoV-2.
* **File 3:** Predicted MHC Class II epitopes from toxins of *Clostridium perfringens*.

The R scripts of Epitope Analyzer can be freely downloaded from **Github** and launched locally.

How to cite Epitope Analyzer

Paper Citation

Contact

Juan Fuxman Bass - fuxman@bu.edu

Luis Soto - lufesu98@gmail.com

**Epitope Distribution section**

Epitope Distribution shows (1) the distribution of the number of unique epitopes that were predicted to bind to selected MHC alleles within an %rank interval and (2) a heatmap showing the presence of epitopes across different proteins. To obtain the distribution of epitopes, users should select one or more MHC alleles, set the min %rank and the max%rank to consider a peptide as an epitope, and select the plot type which could be a histogram or cumulative histogram. The histogram and cumulative histogram plots show the %rank intervals on the x-axis and the number of epitopes on the y-axis. The density is represented with a red line in the histogram plot.  
To show epitopes present across multiple proteins, users need to indicate a minimum number of proteins in which the epitopes are contained. The Shiny app will show a heatmap where the epitopes are on the x-axis while proteins are on the y axis. The cells are colored blue or white indicating the presence or absence of the epitope, respectively.

Parameters

* Condition over MHC alleles: When multiple alleles are selected, users must indicate if they are interested in shared epitopes that bind all the selected MHC alleles (AND) or epitopes that bind at least one of the MHC alleles selected (OR).
* MHC alleles: List of MHC alleles obtained from the input file. Users can choose more than one allele. (Default = first allele)
* Min % rank: Minimum %rank to filter the epitopes. (Default = 0)
* Max % rank: Maximum %rank to consider a peptide as predicted epitope. (Default MHC Class I = 2, MHC Class II = 10).
* Step: Width of the bins in the histogram plot. (Default = 0.1)
* Plot type: The distribution of epitopes can be shown as a histogram or a Cumulative histogram. If a histogram is selected, the density is represented with a red line.
* Min number of proteins: Minimum number of proteins where epitopes are conserved (Default = 2).

**Epitope Density section**Epitope Density shows a scatter plot where the protein length in amino acids is on the x-axis and the number of epitopes is on the y-axis. Hovering over each point shows the name of the protein, number of epitopes, length of the protein, and the epitope density. The epitope density is calculated as the number of epitopes divided by the length of the proteins. It also shows a heatmap representing the number (or density) of epitopes in each protein across each allele.

To obtain the scatter plot, users must indicate the maximum cutoff %rank to consider a peptide as an epitope and click on the “Run analysis” button. This will produce a scatter plot and a table containing the ID of the proteins, their length in amino acids, the number of epitopes, and the epitope density. Clicking on any row of the table will highlight the point in the scatter plot.

To obtain the heatmap, in addition to the maximum cutoff%rank, users must select the plot-type (heatmap or bar plot) and the fill-type (by number or density of epitopes). We recommend using the plot-type “heatmap” when several proteins are analyzed. When heatmap is selected, this shows the alleles on the x-axis and the proteins on the y-axis. The color intensity indicates the log 10 of the number or the density of epitopes. When “barplot” is selected, each protein is represented as a bar while alleles are on the x-axis and the number (or density) of epitopes are on the y-axis. Hovering over each bar (or cell in the heatmap) will show the protein ID, the allele, the number of epitopes, the length of the protein, and the density of epitopes.

Parameters

* Cutoff %rank: Maximum %rank to consider a peptide as predicted epitope. (Default MHC Class I = 2, MHC Class II = 10)
* Color by: Fill heatmap by number or density of epitopes.
* Plot-type: Heatmap is recommended for several proteins, while bar plot is recommended for few proteins.

**Epitope Location section**

Epitope Location shows the location of epitopes within the protein (blue bar). Epitopes are represented on a color gradient from yellow to red reflecting the number of MHC alleles they are predicted to bind to. Hovering on each epitope will show the amino acid sequence, the position within the protein, and the MHC alleles that are predicted to recognize it. Users must select the protein, set a cutoff %rank to filter epitopes, and indicate the MHC alleles of interest. When multiple alleles are selected, users also need to specify if they are interested in epitopes that are predicted to bind all the selected alleles (AND) or at least one of the selected alleles (OR). This tool also shows a table containing the sequence of the epitope, the start and end amino acid position within the protein, the alleles, and the number of alleles to which they were predicted to bind. This analysis may take a few minutes depending on the number of epitopes and the length of the protein.

Parameters

* Protein: List of proteins obtained from the input file. Users must select a protein to be analyzed. (Default = first protein)
* Cutoff %rank: Maximum %rank to consider a peptide as predicted epitope. (Default MHC Class I = 2, MHC Class II = 10)
* Condition over MHC alleles: When multiple alleles are selected, users must indicate if they are interested in shared epitopes that bind all the selected MHC alleles (AND) or epitopes that bind at least one of the MHC alleles selected (OR).
* MHC alleles: List of MHC alleles obtained from the input file. Users can choose more than one allele. (Default = first allele)

**Epitope Promiscuity section**Epitope Promiscuity shows epitopes predicted to bind a minimum number of MHC alleles. The tool depicts a heatmap where alleles are on the x-axis and epitopes are on the y-axis. Users should set the strong and weak binding cutoff %rank to filter the strong and weak epitopes. Strong epitopes are the epitopes with a %rank less or equal than the strong binding cutoff %rank. Weak epitopes are the epitopes with a %rank greater than the strong binding cutoff %rank but less or equal than the weak binding cutoff %rank. In the heatmap, strong epitopes and weak epitopes are colored in red and orange, respectively. Users also need to select a minimum number of MHC alleles to filter and show only epitopes that are predicted to bind to at least this number of alleles. This tool also shows a table containing the amino acid sequence, the position within the protein, the ID of the protein, and the number of alleles to which they are predicted to bind.

Parameters

* Minimum number of alleles: This indicates the minimum number of MHC alleles that an epitope needs to bind to be shown in the heatmap.
* Strong Binding Cutoff %rank: Epitopes with a %rank lower or equal than this cutoff are considered as strong binder epitopes (SB).
* Weak Binding Cutoff %rank: Epitopes with a %rank lower or equal than this cutoff but greater than Strong Binding Cutoff are considered weak binder epitopes (WB).

**Epitope Intersection section**

Epitope Intersection shows an Up-Set plot where the MHC alleles are represented as the sets and the epitopes are represented as the elements. The selected MHC alleles are represented with bars on the left side indicating the number of predicted epitopes for each MHC allele. The number of epitopes for each MHC allele or intersection of MHC alleles is represented with bars at the top. Individual points in the grid indicate epitopes binding to a specific MHC allele while connected points indicate epitopes that can bind to multiple MHC alleles. This tool also provides a table containing the number of alleles, the alleles, the epitope sequences within each region, and their respective number of epitopes.

Parameters

* Cutoff %rank: Maximum %rank to consider a peptide as predicted epitope. (Default MHC Class I = 2, MHC Class II = 10)
* MHC alleles: List of MHC alleles obtained from the input file. Users can choose more than one allele. (Default = first 5 alleles)