Nick Kullman, Graham Clenaghan, Wayne Yang - Data Visualization Spring 2015

Statistical Interactive Explorer of Vaccine Efficacy

Problem

The primary goal of this project is to build a visualization tool to support a technique known as sieve analysis of peptide sequence data from vaccine clinical trials where the vaccine demonstrates partial efficacy.

Sieve Analysis

Sieve analysis aims to compare the protein sequences of the viral strains which are found in the placebo group against the protein sequences of the viral strains found in the vaccine group using a variety of statistical methods [1]. The vaccine can be thought of as a barrier which filters the viral strains observed in the vaccine group.

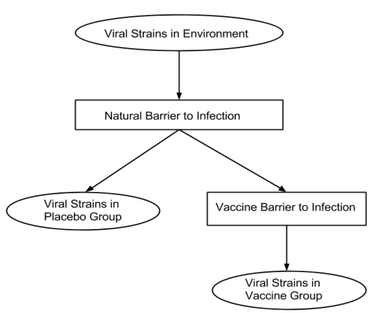


Figure 1

Prior Work

Our project aimed to extend visualizations from the paper [1], in particular figure 2 and figure 3 shown below:

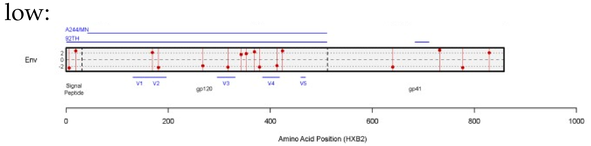


Figure 2: from [1]

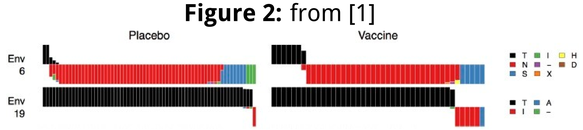


Figure 3: from [1]

Figure 2 shows an overview of the genome for the Env protein with specific sites annotated, and figure 3 shows the distribution of mismatches at two specific sites. The figures were generated with a python script which was run after a statistical analysis determined the sites of interest. We hoped to speed the analysis process through interactivity and then once sites of interest are found, generate usable graphics automatically.

Motivation

AIDS and other viral diseases such as influenza and dengue claim hundreds of thousands of lives each year. To assess the efficacy of vaccines built to combat these diseases, researchers apply statistical methods to the results of clinical trial. Sieve analysis is one such method.

Sieve analysis requires determining which sites in the peptide sequence show statistically significant differences in amino acid expression (mutations) between vaccine and placebo groups. Finding these sites is a labor-intensive process. The effort involved in scanning for important sites in the sequence increases the likelihood of overlooking them. A better tool is needed to increase the speed and ease of browsing peptide sequences and the vaccine- vs. placebo-group differences that exist at each site. Providing this tool will improve the assessment of vaccine efficacy, resulting in quicker turnaround times between vaccine deployments, more effective vaccines, and fewer lives lost.

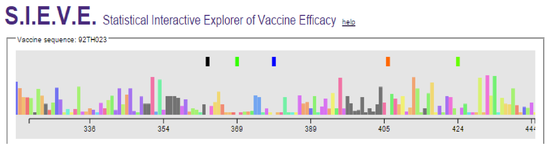
Approach

Our goals in designing the interface were the following:

- Sequence viewing software should be somewhat familiar to researchers.
- Sites of potential interest should be quickly identifiable.
- Sites of known interest should be easily found.

To accomplish this, we decided to break the tool into three parts: an overview showing a summary of the entire data and the ability to drill down and select particular parts, and then once a selection is made, both a summary of the overall statistics and relevant charts for each selected site.

Selection



- Each individual bar corresponds to a particular amino acid.
- Bar height corresponds with a chosen measure of "interestingness."
- Zooming in/out facilitates navigation.

Site-wise Comparisons

- Stacked bar charts are a common visualization for displaying mutations from a reference sequence.
- Different color schemes provide various aggregations of amino acids (based on chemical properties).
- Charts are exportable as SVG files.

Group Data

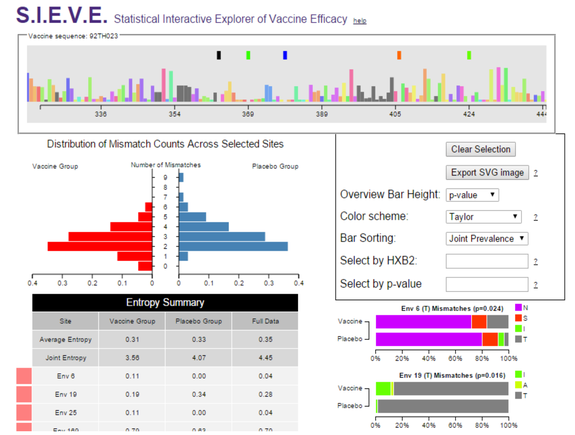
- Pyramid plot provides a view of the distribution of mismatches for each group.
- Table provides both aggregate and site-wise information as well as interactive elements for navigation.

Tools

We chose to accomplish this using the D3.js visualization library so that our web-based interface to be as accessible as possible to researchers in the field. In addition, planned features to aid in editing and exporting the graphic for inclusion in sieve analysis papers and we hoped to design it be easily extensible by researchers to add new features as they saw fit, such as different statistical measures and different data sets.

Results

Final Product



The positioning of the three main sections are based loosely on existing software with a site selection tool on the top of the page with summary and drilldown information available below.

Future Work

- Make the tool more broadly applicable and reduce restrictions on user input.
- Add ability for researchers to share their findings.

References

[1] Paul T Edlefsen, Morgane Rolland, Tomer Hertz, Sodsai Tovanabutra, Andrew J Gartland, Allan C deCamp, Craig A Magaret, Hasan Ahmed, Raphael Gottardo, Michal Juraska, et al. Comprehensive sieve analysis of breakthrough hiv-1 sequences in the rv144 vaccine efficacy trial. *AIDS research and human retroviruses*, 30(S1):A25–A26, 2014.