A Shiny App for Epitope Binning Analyses: High Throughput Analyses Made Easy

Volha Tryputsen[1], Jocelyn Sendecki[2], Robin Ernst[3], Sanjib Dutta[3], Joseph Bourghol[3], Annmarie Winkis[3], Fang Yi[3]

[1]Janssen R&D, MTASS/TMEDS, Raritan, NJ; [2]Janssen R&D, MTASS/TMEDS, Spring House, PA; [3]Janssen R&D, Biologics Research, Spring House, PA

ABSTRACT

Background: Characterizing antibodies through epitope binning is one of the most sought after goals in early research and discovery in antibody therapeutics. The precise epitope that a mAb targets on its antigen can largely dictate its functional activity and help understand the target biology. Recent advancement in label-free biosensor platforms has enabled high throughput epitope screening of large number of mAbs. However, the intrinsic complexity and wealth of information in a 96X96 mAbs pairwise epitope binning experiment call for a deep and better understanding of Wasatch binning software analyses to enable meaningful interpretation of the binning results. A web-based interactive Shiny Epitope Clustering application has been developed based on a collaboration between Translational Medicine and Early Development Statistics (TMEDS), the Structural Biology group and a high-throughput biophysical characterization group within the Hit-to-Lead group to evaluate Wasatch epitope binning software and to reinforce epitope binning analyses and decision-making with statistical support.

Methods: The user starts with a heatmap of pair-wise mAb competition data, saved as a .csv file. In this example the data comes from being processed in Wasatch Binning software but other sources such as IBIS work as well. The Shiny app regenerates the heatmap as well as row-specific, column-specific, and combined dendrograms to visualize initial relationships; provides diagnostic tools to assist users in determining a mathematically-determined number of epitope clusters; and enables visualization and exploration of clusters in 2-dimensional space. Future versions are in development to provide summary reports and potentially visualization in 3-dimensional space.

Results: Epitope Clustering Shiny app functionality helps scientists not only understand and visualize the competitive relationships in the data and make mathematically-based decision about the optimal number of clusters, but also strengthens discovery data integrity (DDI) compliance. Its capacity to analyze combined datasets from different runs could eliminate the need to do multiple overlapping binning experiments for large mAb panels (>96 mAbs). This app is expected to continue to facilitate future high throughput epitope binning analyses and anti-IDs discovery.

OBJECTIVE

The **objective** of this project was to evaluate Wasatch software for epitope binning and to develop analytical tool which enables optimality (mathematical robustness, efficiency of use, web-based functionality, documentation and report generation) and consistency of epitope binning analyses.

METHODS

Data Import

- Clean normalized data in Wasatch Binning software (e.g., exclude inactive mAbs or mAbs that do not self block)
- Export heatmap from Wasatch Binning software
- Resave the heatmap as a .csv file into Excel
- Import .csv file into the app

Clustering

- The heatmap matrix is dichotomized into 0s and 1s based on the threshold values chosen by the user within the app
- The dichotomized heatmap is use to calculate average Euclidean distances for each mAb for rows (ligands), columns (analytes), and for both combined.
- Respective dendrograms are constructed using the Euclidean distances to illustrate hierarchical clustering results

Determining optimal number of clusters

Several statistical methods are offered within the app to augment the user's ability to make a decision about the optimal number of cluster in a data set:

1. Elbow method – plots within groups sum of squares as a function of the number of clusters
2. Silhouette method – plots average silhouette as a function of number of clusters. The silhouette of a data instance is a measure of how closely it is matched to data within its cluster and how loosely it is matched to data of the neighboring cluster, i.e. the cluster whose average distance from the datum is lowest.
3. Bootstrapped AU (Approximately Unbiased)/BP (Bootstrapped Probability) assesses clustering uncertainty using random sampling with replacement.

Clusters visualization

Two methods of cluster visualization are provided. As in the Wasatch binning software, rectangles are overlaid on mAb dendrograms to visualize groups. Additionally, Principal Components Analysis (PCA) is used to visualize mAbs in the dimension of the first two principal components, color-coded by the number of clusters selected by the user, who is informed by the number of clusters suggested by the elbow, silhouette, and AU/BP methods.

Network generation

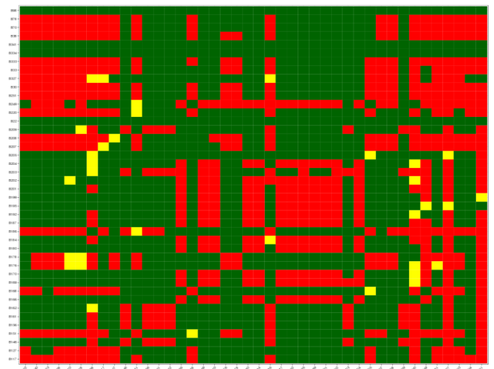
Two sets of network plots are available; both use the following Community Network Algorithms:

1. **Walktrap**: An agglomerative approach based on random walks. Walktrap runs short random walks 4 steps and uses the results of these random walks to merge separate communities.
2. **Multilevel Community**: Communities are not merged as in other methods; rather, nodes are moved between communities such that each node makes a local decision that maximizes its own contribution to the modularity score.
3. **Leading Eigenvector**: is a top-down hierarchical approach that optimizes the modularity function. In each step, the graph is split into two parts. The split is determined by evaluating the leading eigenvector of the so-called modularity matrix.

RESULTS

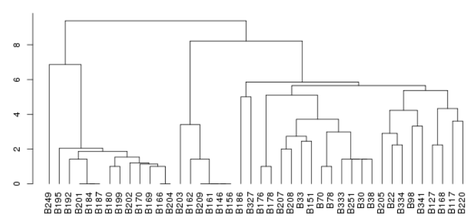
Imported from Wasatch Binning software data is visualized in a heatmap (**Figure 1**) and color-coded by red (if mAb binds to the same epitope), yellow (if value is just above the threshold set for binding) or green (mAb binds to different epitope)

Figure 1. Heatmap



Combined binary dendrogram (**Figure 2**) Items that are "closest" to each other using a distance metric are grouped together into small groups; these small groups are merged into larger groups until all the items are grouped together at the top of the dendrogram

Figure 2. Combined binary dendrogram



When choosing an optimal number of clusters one should strike for the balance between maximum compression of the data to a single cluster, and maximum accuracy by assigning each data point to its own cluster.

Figure 3 (left) shows an Elbow Plot which suggests to choose number of clusters indicated by bend at "elbow" or circle size (the larger the better).

Figure 3 (right) shows a Silhouette Plots and suggests to choose the number of clusters, corresponding to the largest average silhouette width, indicated by the dashed vertical line.

Figure 3. Elbow Plot (left) Silhouette Plot (right)

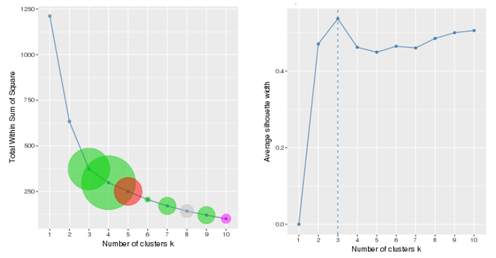


Figure 4 highlights the clusters which could stably be observed if we increase the number of observations, i.e., if the same experiment was repeated hundreds of times, these are the clusters that would reliably emerge.

Figure 4. Bootstrapped AU/BP Plot

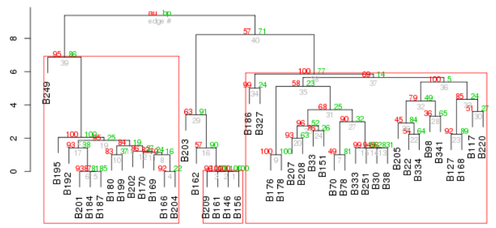
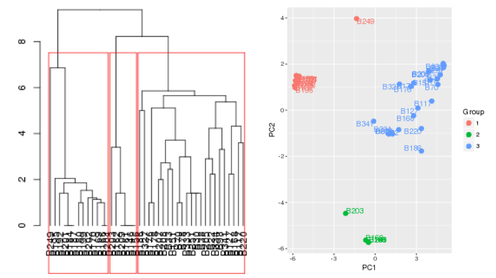


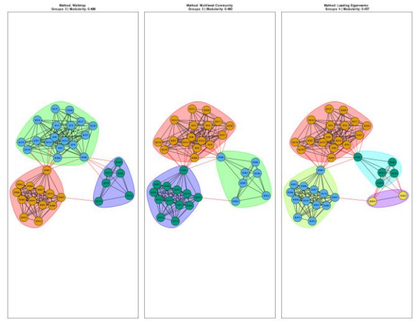
Figure 5 enables mAb grouping visualization using both a dendrogram and a plot of the first two principal components. By applying different number of clusters (here the chosen number of clusters is 3) a user has an opportunity to further explore the data and to make a final decision on the optimal number of clusters.

Figure 5. Hierarchical clustering and PCA by the number of clusters



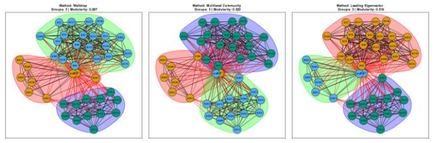
Network plots using combined dichotomized heatmap information with three different Community Network Algorithms is shown in **Figure 6**.

Figure 6. Network plots using combined transposed heatmap data



Network plots using heatmap data as-is with three different Community Network Algorithms is shown in **Figure 7**.

Figure 7. Network plots using heatmap as-is



To decide on the number of groups a user has to evaluate the magnitude of the modularity score (a metric of the strength of the division of a network into groups) and choose the results of the method with the highest modularity number.

CONCLUSION

- Epitope Clustering Shiny App serves as an extension of Wasatch Binning software for epitope binning analyses and provides custom-made data exploration and visualization tools
- By enabling objective mathematically driven decisions about epitope binning analyses, Shiny App strengthens DDI compliance.
- Shiny App capacity to analyze combined datasets from different runs could reduce the need for multiple overlapping binning experiments for large mAb panels.

REFERENCES

1. Wasatch Microfluidics. Array SPR Applications: Epitope Binning. White paper
2. Noah T. Ditto and Ben D. Brooks. The emerging role of biosensor-based epitope binning and mapping in antibody-based drug discovery. Expert Opinion On Drug Discovery, Vol. 11 , Issue 10,2016