Cluster analysis is an important data mining technique that is widely used in biology. Combined with visualization, it is employed to partition large datasets into meaningful subsets that can reveal important biological relationships. Frequently, clustering is used to analyze gene-by-sample expression matrices and can help to identify disease subtypes as well as sets of co-expressed or functionally related genes.

Clustering, however, is not perfect and its results depend on a many factors. Common challenges include the selection of algorithm, parameters, and the similarity metric used. Therefore, it is important to enable analysts to evaluate the quality of a particular clustering and to compare clustering results. However, existing visualization tools that allow cluster comparison only show high-level relationships of clusters.

To enable analysts to examine low-level relationships and make informed decisions about cluster refinement, we have developed visualization tools to interactively explore and evaluate cluster assignments in both discrete and fuzzy clustering results. Analysts are able to run various clustering algorithms with a selection of similarity metrics, visually examine the specificity or uniqueness of individual elements, split ambiguous clusters by specifying thresholds, and manually refine cluster assignments. We supplement heatmap views with a visualization of the cluster fit for both the cluster that an element is assigned to and for all other clusters. We embedded our tools into Caleydo StratomeX (http://stratomex.caleydo.org) to enable comparison of stratifications based on clusterings or categorical data. StratomeX can integrate multiple molecular and clinical data types to support the resolution of cluster assignments.