The Tunable Biclustering Algorithm (TuBA) is a graph-based biclustering algorithm that identifies co-expressed gene sets within subsets of samples. Underlying TuBA is a simple yet novel pairwise proximity measure - given a pair of genes, samples that constitute the top (or bottom) percentile sets for the respective genes are shortlisted. Next, the overlap between their percentile sets is examined to determine if it exceeds the user specified threshold. Gene-pairs that exceed the threshold reflect a co-expression signature of potential interest. We represent such gene-pairs graphically as a pair of nodes linked by an edge, where the edge corresponds to the set of samples that are shared between the percentile sets of the two genes. Subsequently, using an iterative approach we identify the biclusters in these graphs.

The “tunable” nature of TuBA pertains to the choice of 2 parameters:

1. The size of the percentile sets
2. The threshold for minimum overlap (based on Jaccard index) between the percentile sets

Based on the choices of these 2 parameters the overall graph and consequently the final biclusters are determined - smaller sizes of percentile set (first parameter) can reveal co-expression signatures that are rarer and present in fewer samples/patients; the values of the Jaccard index threshold for overlap between percentile sets determines the confidence about the co-expression of genes identified in the biclusters with larger thresholds associated with greater confidence