

The functions of lncRNAs are mediated by intermolecular interactions. Detailed mapping of interactions through laboratory experimentation is limited by reagent availability and other resources. To overcome these limitations, machine learning approaches can be used to predict interactions between RNAs and the proteins that interact with them for regulation of the RNAs, and/or mediation of their functions. To determine the most optimal set of methods that would output the accurate protein pairs, seven filters and two correlation techniques were tested. The filters investigated include the Bins 1 Filter, Bins 2 Filter, Variance Filter, Smooth Filter, Cutoff Filter, and Outlier Filter. The two correlations techniques tested include the Euclidean Distance and Pearson Correlation. Significantly correlated protein pairs following filtering methods applied to the binding probabilities were found to match up with outputted graphs, showing veritability of the methods. The highly correlated protein pair outputted using the Pearson correlation has greater similarity than the highly correlated protein pair outputted using Euclidean distance. An analysis of the binding probabilities following filtering reveal that the Smooth filter accentuates the extreme values the most. Background research needs to be conducted on each protein pair to determine if there is a significant correlation in its function.