

Logic For Candidate Class:

There are __ criteria that need to be considered based on Brown et al. These are if it is a transcript is classified as coming from a miRNA gene, and the foldback structure of the gene which encodes the candidate sequence, the majority of sequences in this sample being miRNA in the csv output file, and the alignments of the sequences being in a bimodal distribution being in on of two regions.

Each of these should be weighted evenly to average out at the end of the pipeline.

Classifying using it coming from an miRNA gene:

Check the first few letters of the gene name for that candidate sequence. If it corresponds to the expected one for miRNA (MIR) then the attribute for classifying as coming from an miRNA gene should be set to 1, otherwise it should remain at 0.

Classifying using the majority of the sequences as being miRNA based on the naming convention:

The values needed are in the csv file written out of the pipeline as rnAmt.csv. Check the percentage of the sequences that are miRNA, if it is above 75% then the attribute for classifying if the majority of the sequences are miRNA should be set to 1, otherwise it should remain at 0.

Classifying using the alignments of the sequences being in a bimodal distribution in one of two regions:

The location of where the alignment is is known. The alignment can be checked for being in each of the regions, and for the bimodal distributions looking at the text files. If they satisfy the region and bimodal distribution then the attribute for this condition is set to 1, otherwise it remains at 0.

Classifying using the foldback structure:

The foldback structure that is expected is the two regions falling across from one another. This can be assessed for looking at the ps file. If this is the case then the attribute for this condition is set to 1, otherwise it remains at 0.

All coming together:

At the end of the pipeline, all of the above attributes can be added up then divided by the number of them (in this case 4). If a candidate reaches all 4 of these criteria then it is confidently classified as miRNA.