1. What problem are you investigating? Why is it interesting?
2. What dataset are you going to use? How are you planning on acquiring the dataset and performing any necessary annotation?
3. How is deep learning going to help you with your science? Is it automating a laborious task, performing inference or some other application? What method or algorithm are you implementing? Your answer does not need to be precise, but it should give us an idea of where you are heading.
4. Is deep learning an appropriate approach to solving this problem? Why?
5. How are you going to evaluate your results? Consider both qualitative and quantitative analysis
6. miRNA-gene target binding prediction. miRNA bind to mRNAs and regulate the protein synthesis of these RNAs. Predicting miRNA binding could be a new way for scientists to modulate protein synthesis for scientific experiments and provide insights to RNA-RNA binding patterns overall.
7. The TDC miRTarBase dataset. The dataset can be downloaded through the TDC package and is overall ready for ML use. The dataset provides the miRNA sequence and the target amino acid sequence, not the mRNA sequence that the miRNA actually binds to. Directly predicting the protein sequence rather than the mRNA sequence eliminates an extra layer of data from redundant wobble bases, which would then require significantly more data if all the possible mRNA sequences are calculated.
8. Automating a laborious task, performing inference. Predict which proteins a miRNA sequence will bind to and which ones it won’t. This is a binary classification task.
9. Yes – the algorithm will be able to extract features from the protein sequence that we would not necessarily predict without generating all the possible RNA sequences from the protein sequence.
10. We can evaluate qualitatively by considering the features extracted and seeing if there is a physical representation/meaning to those features. We can quantitatively analyze the data by doing tests for accuracy and precision.