ML and DL for ncRNA prediction and classification problems:

Deep learning (DL) models are growing in popularity for ncRNA prediction and classification problems.
https://doi.org/10.1371/journal.pcbi.1008415
DL combined with ncRNA motif finding:

- Use pre-trained classification and function prediction models to complement the ncRNA discovery pipeline from (Yao et al., 2007)
- Some suggested steps:
  - Review state-of-the-art DL (or ML) models for short ncRNA function prediction/classification models
  - For discovered motifs, create test sets of your motif example sequences and predict them

Potential interesting questions:

- **Feature attribution**: figuring out what inputs were the most important for a model’s prediction.
- For sequence based ncRNA DL predictors, how do their predictions relate to RNA secondary structures? Can we apply current feature attribution methods to figure this out?

[https://github.com/slundberg/shap](https://github.com/slundberg/shap)
● Some suggested steps:
  ○ Review state of the art DL models for those which use only sequence features for predictions & select some to try
  ○ Review state of the art feature attribution methods for DL networks (deepSHAP, deepLIFT, Integrated Gradients, etc.) & select some to try
  ○ Gather a test set of structures and sequences
  ○ Predict the function/families and run the attribution methods for these predictions
  ○ Visualize importance
  ○ Look for what the networks might learn implicitly about secondary structure