Supporting material for the RECOMB 2011 paper

Formal and complete descriptions of the various feature representations compared in the paper is available [here](#) (this material is included in the JCB version of the paper).

Datasets

All datasets (S-Full, S-Full-Test, S-Full-Train, and S-Strand2) are taken from the paper *Computational approaches for RNA energy parameter estimation*, Mirela Andronescu, Anne Condon, Holger H. Hoos, David H. Mathews and Kevin P. Murphy. RNA 2010. 16: 2304-2318.

Additional trained scoring models

RNA-family specialized scoring models are available [here](#). These models were obtained by training the StHighCoHigh model on S-Full, and then re-training it separately on the members of each one of the main RNA families in S-Full. Thus, these models are over-fitted to the corresponding families, where a family-model is expected to predict better inputs which belong to its family.

Predictions on S-Full-Test and S-Strand2

Predictions of the models described in the paper on the S-Full-Test and S-Strand2 datasets. All these predictions were obtained by models trained on the S-Full-Train dataset, where features which absolute weight is lower than 0.01 fraction of the maximum absolute feature weight were filtered (corresponding to Table 3 in the RECOMB paper). First bracket line is the real structure, and second line is the predicted one. **Note that S-Strand2 overlaps to some extent with the training set.**

- model St\textsuperscript{med}Co\textsuperscript{med}
  - Predictions on S-Full-Test.
  - Predictions on S-Strand2.
- model St\textsuperscript{high}Co\textsuperscript{med}
  - Predictions on S-Full-Test.
  - Predictions on S-Strand2.
- model St\textsuperscript{med}Co\textsuperscript{high}
  - Predictions on S-Full-Test.
  - Predictions on S-Strand2.
- model St\textsuperscript{high}Co\textsuperscript{high}
  - Predictions on S-Full-Test.
  - Predictions on S-Strand2.