

Prediction of RNA and DNA binding sites: weekly report

Pandu Raharja ^{1,2} Julian Schmidt ^{1,2}

¹Technische Universität München

²Ludwig-Maximilians-Universität München

November 17, 2016

Outline

- ▶ What we did.
- ▶ Problems.

Repository and Datasets

- ▶ Scripts and results could be found in <https://github.com/raharjaliu/BIFers> (except for huge files)
- ▶ **To Jonas:** input data for `profkernel` could be found at `/mnt/project/pp2_1617/xna_raharjaschmidt`.

Stats

Results of redundancy reduction on training and negative sets:

Category	numbers
swissprot_all	552884
negative_all	539298
negative_all_cdhit (default params: 90% ID)	317077
negative_all_cdhit (80% ID)	256898
negative_all_cdhit (70% ID)	207737
negative_human	20005
negative_human_reduced (hval: 0)	1813
negative_all_reduced (hval: 0)	still running...
target_dna	264
target_rna	442
target_dna_reduced (hval: 20)	213
target_rna_reduced (hval: 20)	358