

# Prediction of RNA and DNA binding sites: weekly report

Pandu Raharja <sup>1,2</sup>    Julian Schmidt <sup>1,2</sup>

<sup>1</sup>Technische Universität München

<sup>2</sup>Ludwig-Maximilians-Universität München

December 15, 2016

# Outline

- ▶ Who's presenting what:
  - ▶ Julian: XNA binding prediction
  - ▶ Pandu: XNA binding site prediction
- ▶ Problems.

As always:

- ▶ Scripts and results could be found in:  
<https://github.com/raharjaliu/BIFers>

# Part I: XNA binding prediction

# Current state

1. Train profkernel: finished (for split 1 and 2)
2. Predict binding: finished (for existing models)
3. Evaluate predictions for different  $k$  and  $s$ : Work in progress

## Part II: XNA binding site prediction

# First thing first

- ▶ Errors are indeed at P17574 and P17574:
  - ▶ Removed P17574 or P24264: `pp2features.py` finishes.
  - ▶ Included P17574 or P24264: `pp2features.py` crashes.
  - ▶ Running **only** P17574 or P24264: `pp2features.py` crashes.
- ▶ Extracted features could be found in  
`/mnt/project/pp2_1516/xrna_raharjaschmidt/  
machine_learning/{dna_big.arff,rna_big.arff}`.

# Machine Learning Sorcery

- ▶ `pp2features.py` exports to csv: py continuation!
  - ▶ jupyter: interactive analysis and reporting
  - ▶ scikit: features selection + basic ML
  - ▶ pandas: data manipulation
  - ▶ etc
- ▶ Analysis runs natively on Python now → more coming soon.



# References