Prediction of RNA and DNA binding sites: weekly report

Pandu Raharja [1,2] Julian Schmidt [1,2]

[1]Technische Universität München

[2]Ludwig-Maximilians-Universität München

December 8, 2016

Outline

- What we did:
  - Julian: xRNA binding prediction
  - Pandu: xRNA binding site prediction
- Problems.

As always:

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

Part I: xRNA binding prediction

Part II: xRNA binding site prediction

Project Specification

- Java is for enterprise solution.
- Development in Python $\rightarrow$ jupyter, scipy, scikit-learn, (Theano, TensorFlow, CNTK et al. if ever needed)

Dataset

- We're a bit late into the game due to conference and illness.
- Extracted features could be found in `/mnt/project/pp2_1516/xrna_raharjaschmidt/machine_learning/{dna_big.arff,rna_big.arff}`

Extraction of Features on ppDNA2

- query.disis not found.
- Q84ZU4: no significant pfam hit, using default values...
- P03206: no significant pfam hit, using default values...

Extraction of Features on ppRNA2

- Processing P17574... ParseError: "It seems that we have a situation now. The expected amount of columns is 22, found: %s!" % len(tokens)
- P24264 **(idem)**
- P67876: no significant pfam hit, using default values...
- P0C206: no significant pfam hit, using default values...
- P0C8P8: no significant pfam hit, using default values...
- P07243: no significant pfam hit, using default values...
- Q57817: no significant pfam hit, using default values...
- P04891: no significant pfam hit, using default values...
- P18683: no significant pfam hit, using default values...

References

- Feigelman, J. (2016). "Stochastic and deterministic methods for the analysis of Nanog dynamics in mouse embryonic stem cells." PhD Thesis, Technische Universität München, Munich, Germany.
- Hoppe, P.S., Schwarzfischer, M., Loeffler, D., Kokkaliaris, K.D., Hilsenbeck, O., Mortz, N., ... & Etzrodt, M. (2016). Early myeloid lineage choice is not initiated by random PU.1 to GATA1 protein ratios. *Nature*, **535(7611)**, 299-302.
- Graf, T., & Enver, T. (2009). Forcing cells to change lineages. *Nature*, **462(7273)**, 587-594.