HMMicro

Predicting miRNA targets from epigenetic data and protein-RNA interactions

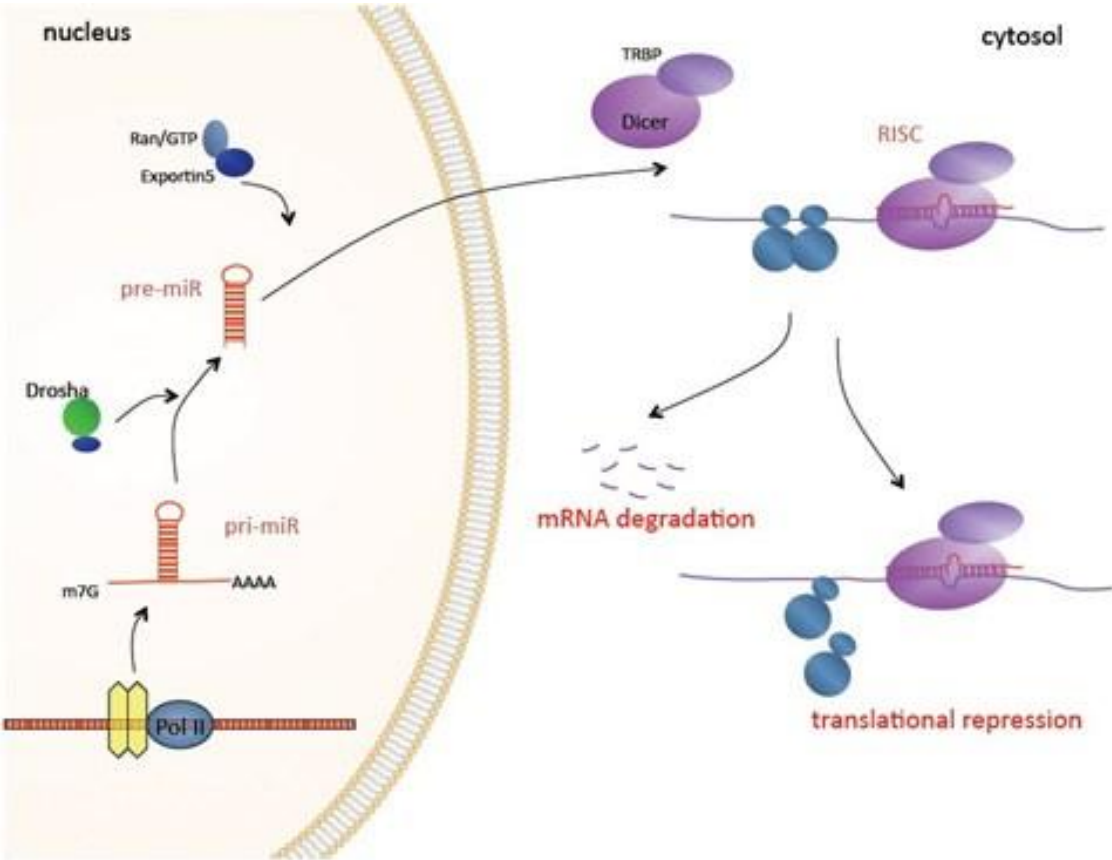Felix Wu and Nicholas Giangreco

Agenda

- Biological overview of micro RNAs (miRNAs)
- The problem: finding miRNA targets and binding sites
- A Hidden Markov Model (HMM) implementation
  - Features data
  - Training/Test data
- Timeline and goals
- Questions

Photo: Thomas Bresson

Introduction to miRNAs

- miRNA: short ~22 nt RNAs
  - Produced from larger RNA molecules
  - Come in "clusters"
- Key gene regulatory component
- Important roles in phenotypes, health, disease
- E.g., development of bilaterian animals
  - *mir-100/let-7/ mir-125*

Problem: Discovering miRNA binding sites

- Motivation:
  - miRNA binding rules are complicated; difficult to experimentally assay targets
  - Gene regulation at any and all levels are dependent; act in a concerted or complementary manner for a desired outcome.
  - Regulatory logic from many sources of epigenetic information
    - chromatin modifications
    - nucleosomal-free regions
    - gene expression
    - DNA methylation
    - etc.
  - Is the regulatory logic from other levels informative?
- How do we approach this problem?
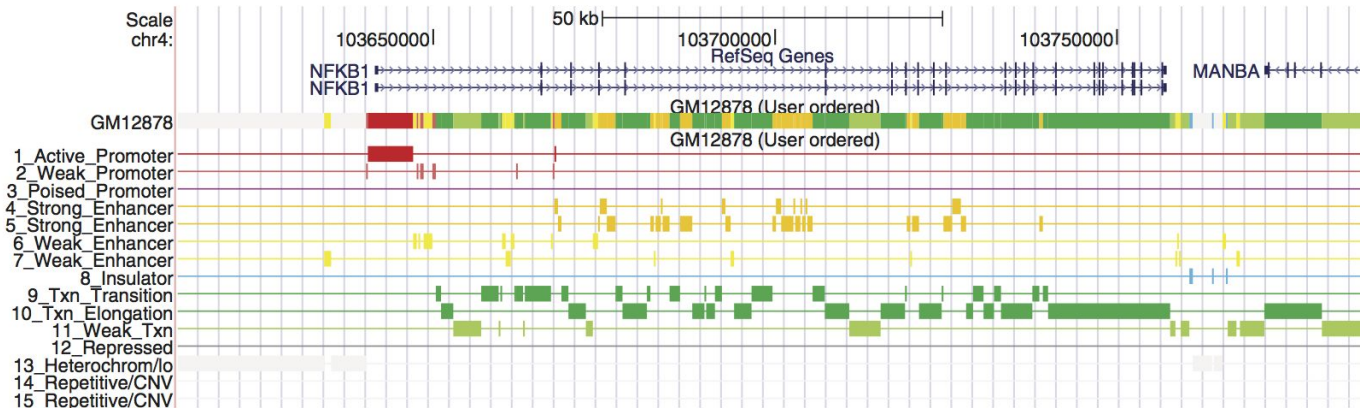 - Hidden Markov Model

HMM for miRNA binding site discovery

- Hidden states
  - Binding vs. non-binding region
  - Cluster region
  - Cluster type?

- Observations
  - Epigenetic features
  - Protein-RNA interactions

Ernst, J., & Kellis, M. (2012). *Nat Methods*. http://doi.org/10.1038/nmeth.1906

Experimentally-found miRNA binding sites (training/test sets)

- Lipchana et al. 2011 *Gen. & Dev.*
  - +20,000 experimentally found miRNA binding sites in human embryonic stem cells (hESCs)
- Separate miRNA binding sites into training and test sets

Potential feature data

- PAR-CLIP RBM10 (RNA-binding protein) from kidney stem cell line HEK293
  - GSM1095142
  - GSM1095143
- Transcription factor, RNA, Histone and Methylation sequencing data from hESCs
  - ENCODE
  - GSE36114
- Looking into other relevant data sets

Tools

- Pomegranate - https://github.com/jmschrei/pomegranate
  - Python library for Hidden Markov Model design
- GitHub
  - Code organization and reproducibility

Timeline

3/6 - 3/13

- **Processing of experimental data sets, establishing HMM states**

3/13 - 3/20

- **Establish a dimension reduction scheme**

3/20-3/27

- **Execute dimension reduction scheme**

3/27 - 4/3

- **Build and learn HMM (Baum-Welch algorithm)**

4/3 : MIDTERM PRESENTATION

4/3 - 4/10

- **Test HMM**

4/10 - 4/17

- **Tweak and test again**

4/17 - 4/24

- **(Incorporate more states, data types)**

4/24 - 5/1

- **Finish modeling and complete presentation**

5/1 : FINAL PRESENTATION

Questions?

References

Encode Consortium. (2012). An integrated encyclopedia of DNA elements in the human genome. *Nature*, *489*(7414), 57–74. http://doi.org/10.1038/nature11247

Ernst, J., & Kellis, M. (2012). ChromHMM: automating chromatin-state discovery and characterization. *Nat Methods.*, *9*(3), 215–6. http://doi.org/10.1038/nmeth.1906

Kim, N., & Ha Minju. (2014). Regulation of microRNA biogenesis. *Nat Rev Mol Cell Bio.*, 509-24. http://www.nature.com/nrm/journal/v15/n8/full/nrm3838.html

Lipchina, I., Elkabetz, Y., Hafner, M., Sheridan, R., Mihailovic, A., Tuschl, T., … Betel, D. (2011). Genome-wide identification of microRNA targets in human ES cells reveals a role for miR-302 in modulating BMP response, *2*, 2173–2186. http://doi.org/10.1101/gad.17221311

Yates, L. A., Norbury, C. J., & Gilbert, R. J. C. (2013). The long and short of microRNA. *Cell*, *153*(3), 516–519. http://doi.org/10.1016/j.cell.2013.04.003