*Bioinformatics*, 2017, 1-2

doi: 10.1093/bioinformatics/boss

Advance Access Publication Date: 08 May 2017

Cool Things

|  |
| --- |
| Cool Things  HMMicro: predicting miRNA targets from epigenetic data  Wu, Felix\*,1 and Nick Giangreco\*,1  1Department of Systems Biology, Columbia University, New York, NY, USA  \*To whom correspondence should be addressed.  Associate Editor: Voldemort  Received on None; revised on of your; accepted on business  Abstract  **Motivation:** Multiple gene regulatory layers give rise to complex phenotypes, which include redundant patterns to maintain homeostasis and other functions. Identifying miRNA binding sites, another regulatory layer at the post-transcriptional level, is difficult and infeasible to obtain for any given condition and treatment.  **Results:** We developed a Hidden Markov Model (HMM) for predicting *de novo* miRNA binding sites in HEK293 cells using epigenetic information provided by ENCODE. Our model shows a low error rate and a high true positive rate, saying that leveraging known epigenetic information of a cell can be useful in predicting novel regulatory layers.  **Availability:** https://github.com/ngiangre/HMMicro  **Contact:** flw88@cumc.columbia.edu  **Supplementary information:** Supplementary data are available at *Bioinformatics* online. |

# Introduction

The central dogma of molecular biology tells us that DNA is transcribed to RNA, RNA is translated to protein, and information cannot not flow from protein back to the previous molecules. The population of proteins give rise to the complex and dynamic cellular phenotype that keeps homeostasis or gives rise to disease. The phenotype of a cell is a product of the reactions and relationships between many molecular layers, such as chromatin modifications, transcription factor binding, and chromatin confirmation. Each layer provides a regulatory logic, which necessarily robust for maintaining homeostasis when bombarded by its environment.

We hypothesize that there exists redundancy between the many layers that produce the population of proteins in a cell. This redundancy is found in molecular patterns that is present in all layers, allowing for robust phenotypes. Thus, integrating knowledge from all different molecular layers can give us a more succinct observation of the underlying phenotype of a cell.

With the advent of next generation sequencing and the popularity of high throughput experimentation, all molecular layers are not able to be assayed due to time and financial constraints. While understanding molecular phenotypes can help elucidate molecular mechanisms in health and disease, predicting patterns in other molecular layers can help in uncovering complex cellular phenotypes.

We present a HMM for *de novo* prediction of miRNA binding sites in HEK293 cell line. We train our model using various epigenetic experiments assayed through the ENCODE consortium, and test our model using experimentally validated miRNA binding sites for this model system. We obtain a low prediction error and high prediction of known miRNA binding sites. This method can be expanded upon for leveraging the vast amounts of existing information that give rise to a cellular phenotype.

# Methods

We obtained HEK293 epigenetic signal

# Results

## Multiple epigenetic layers contain redundant information

### 3.1.1 This is heading 3 style

**Fig. 1. Relation between τ and *t*.** This example has only two continuous Steppers, S1 and S2.

**Table 1.**Benchmark results of the cascade oscillators model

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| |S| | Predicted cost | Timing | Predicted speed | Speed |
| 1 | S219.20(100%) | 68m43s | 1.00 | 1.00 |
| 2 | 29.10+219.10(~50%) | 35m13s | 2.00 | 1.95 |
| 4 | 219.20(100%) | 68m43s | 1.00 | 1.00 |
| 10 | 29.10+219.10(~50%) | 35m13s | 2.00 | 1.95 |
| 20 | 219.20(100%) | 68m43s | 1.00 | 9.5 |

This is table foot note sample text This is table foot note sample text This is table foot note sample text

Acknowledgements

We’d like to thank Itsik Pe’er and the Spring 2017 Computational Genomics class at Columbia University for feedback and support.

Funding

This work has been supported by the training grant from the Integrated Program in Biomedical, Cellular, and Molecular Sciences at Columbia University Medical Center.

*Conflict of Interest:* none declared.

References

Hafner, M. (2010) Transcriptome-wide identification of RNA-binding protein and microRNA target sites by PAR-CLIP. *Cell*, 141(1): 129-41.