Modelling complex features from histone modification signatures using genetic algorithm for the prediction of enhancer region

Article *in* Bio-medical materials and engineering · September 2014

DOI: 10.3233/BME-141210 · Source: PubMed

3 authors, including:

Nung Kion Lee
University Malaysia Sarawak
49 PUBLICATIONS **145** CITATIONS

SEE PROFILE

Mohd Tajuddin Abdullah
Universiti Malaysia Terengganu
437 PUBLICATIONS **1,079** CITATIONS

SEE PROFILE

Bio-Medical Materials and Engineering 24 (2014) 3807–3814
DOI 10.3233/BME-141210
IOS Press

Modelling complex features from histone modification signatures using genetic algorithm for the prediction of enhancer region

Nung Kion Lee[a,*], Pui Kwan Fong[a], and Mohd Tajuddin Abdullah[b]
[a]*Faculty of Cognitive Sciences and Human Development, Universiti Malaysia Sarawak, Kota Samarahan, Malaysia*
[b]*Center of Tasik Kenyir Ecosystem, Universiti Malaysia Terengganu, Kuala Terengganu, Malaysia*

Abstract. Using Genetic Algorithm, this paper presents a modelling method to generate novel logical-based features from DNA sequences enriched with H3K4mel histone signatures. Current histone signature is mostly represented using k-mers content features incapable of representing all the possible complex interactions of various DNA segments. The main contributions are, among others: (a) demonstrating that there are complex interactions among sequence segments in the histone regions; (b) developing a parse tree representation of the logical complex features. The proposed novel feature is compared to the k-mers content features using datasets from the mouse (mm9) genome. Evaluation results show that the new feature improves the prediction performance as shown by f-measure for all datasets tested. Also, it is discovered that tree-based features generated from a single chromosome can be generalized to predict histone marks in other chromosomes not used in the training. These findings have a great impact on feature design considerations for histone signatures as well as other classifier design features.

Keywords: Genetic algorithm, tree-based feature, histone feature

1. Introduction

Comprehension of gene regulation involves locating the cis-acting regulatory elements comprising clusters of transcription factor binding sites (TFBS) that initiate the mechanism of gene transcription. Enhancers are a type of cis-regulatory element that promote gene expression and often are essential for eliciting complex expression patterns of developmental genes. An enhancer region typically spans a few hundred base pairs (bp) comprising clusters of TFBSs (at multiple sites) that work in cis--each site is about 6 to 20bp in length.

The first two authors contributed equally to this paper.

*Corresponding author: Nung Kion Lee, Faculty of Cognitive Sciences and Human Development, Universiti Malaysia Sarawak, Kota Samarahan, Sarawak, Malaysia. E-mail: nklee@fcs.unimas.my, Tel:+6082-584152, Fax:+6082-581567