



## Review Article

## A survey of recently emerged genome-wide computational enhancer predictor tools

Leonard Whye Kit Lim<sup>a,\*</sup>, Hung Hui Chung<sup>a,\*</sup>, Yee Ling Chong<sup>b</sup>, Nung Kion Lee<sup>c</sup><sup>a</sup> Animal Biotechnology Lab II, Department of Molecular Biology, Faculty of Resource Science and Technology, Universiti Malaysia Sarawak, 94300 Sarawak, Malaysia<sup>b</sup> Department of Zoology, Faculty of Resource Science and Technology, Universiti Malaysia Sarawak, 94300 Sarawak, Malaysia<sup>c</sup> Department of Cognitive Sciences, Faculty of Cognitive Sciences and Human Development, Universiti Malaysia Sarawak, 94300 Sarawak, Malaysia

## ARTICLE INFO

## Article history:

Received 13 December 2017

Received in revised form 13 March 2018

Accepted 13 March 2018

Available online 16 March 2018

## Keywords:

Enhancer prediction

Supervised learning

Unsupervised learning

Semi-supervised learning

## ABSTRACT

The race for the discovery of enhancers at a genome-wide scale has been on since the commencement of next generation sequencing decades after the discovery of the first enhancer, SV40. A few enhancer-predicting features such as chromatin feature, histone modifications and sequence feature had been implemented with varying success rates. However, to date, there is no consensus yet on the single enhancer marker that can be employed to ultimately distinguish and uncover enhancers from the enormous genomic regions. Many supervised, unsupervised and semi-supervised computational approaches had emerged to complement and facilitate experimental approaches in enhancer discovery. In this review, we placed our focus on the recently emerged enhancer predictor tools that work on general enhancer features such as sequences, chromatin states and histone modifications, eRNA and of multiple feature approach. Comparisons of their prediction methods and outcomes were done across their functionally similar counterparts. We provide some recommendations and insights for future development of more comprehensive and robust tools.

© 2018 Elsevier Ltd. All rights reserved.

## Contents

1. Introduction	132
2. Supervised machine learning	133
2.1. Sequence feature	133
2.2. Chromatin states	136
2.3. Multiple feature approach	136
3. Unsupervised machine learning	137
3.1. eRNAs	137
3.2. DNaseI hypersensitivity site (DHS)	138
3.3. Multiple feature approach	138
4. Semi-supervised machine learning (co-training)	138
5. Regulatory variants prediction	139
6. Comparative genomics	139
6.1. The future of enhancer predictor tool	139
References	140

## 1. Introduction

Our perspective on non-coding regions in the genome had been overturned since King and Wilson (1975) discovered that the non-coding regions are one of the major players in divergent evolution in the animal kingdom almost forty years ago. One of the most

\* Corresponding authors.

E-mail addresses: [16020176@siswa.unimas.my](mailto:16020176@siswa.unimas.my) (L.W.K. Lim), [hhchung@unimas.my](mailto:hhchung@unimas.my) (H.H. Chung).