Yale University

MB&B
260/266 Whitney Avenue
PO Box 208114
New Haven, CT 06520-8114

Telephone:
203 432 6105
360 838 7861 (fax)
mark@gersteinlab.org
www.gersteinlab.org

June 14, 2016

Dear Dr. Rusk,

Please find our enclosed manuscript entitled "Supervised enhancer prediction with epigenetic pattern recognition and targeted validation across organism", which we hope will be considered for publication in Nature Methods. I personally talked about this study with you while you were visiting Yale. In this method, we aggregated the epigenetic signals from large scale of validated enhancers, extracted patterns of features in a supervised fashion, and scan the whole genome with matched filter. To the best of our knowledge, it is the first endeavor to apply signal processing methods to analyze epigenetic signals for enhancer prediction. The dominating peak-trough-peak pattern that is observed within the signal of certain post-translational histone modifications at active enhancers has natural corresponding biological mechanisms. The model also aggregates different epigenetic in a cell-type dependent fashion, which allows the model to be applied to many different cell lines and species. Our method is validated through multiple rounds of *in vivo* and *in vitro* assays across species.

The new method we developed is trained with the output of massively parallel reporter assays. Traditionally, enhancers were characterized using low throughput validation assays, resulting in rigorous validation of very few cell-type specific mammalian enhancers. Those enhancers were also typically selected based on certain genomic characteristics, which introduces selection bias. Both the experimental size and selection biasness hindered effective training and cross-validation of enhancer prediction models. The development of a large number of massively parallel reporter assays allows for the first time the identification of thousands of enhancers. Using data from these assays, we are able to rigorously train and test statistical models for enhancer prediction. The large number of validation assays we performed in mammalian cell lines and tissues also provides useful resource to the community.

We believe that this new tool will be useful to researchers. Its ability to predict tissue-specific enhancers across species allows it for broad application. The source code of the software is readily available in the github repository (https://github.com/gersteinlab/MatchedFilter). We are already using the method to predict enhancers in a few publications resulting from the third phase of the ENCODE consortium. Recent experimental validations also showed that the enhancers predicted by our method is capable of regulating targeted genes in the reporter assay.

We have made significant changes based on feedbacks and suggestions from our last submission. We hope you would reconsider it for publication in Nature Methods.

We have listed a number of suitable reviewers for this work.

Yours sincerely,
Mark Gerstein
Albert L. Williams Professor
of Biomedical Informatics