

ORAL PRESENTATION

Open Access

A hidden Markov model for detecting multi-gene chromatin domains

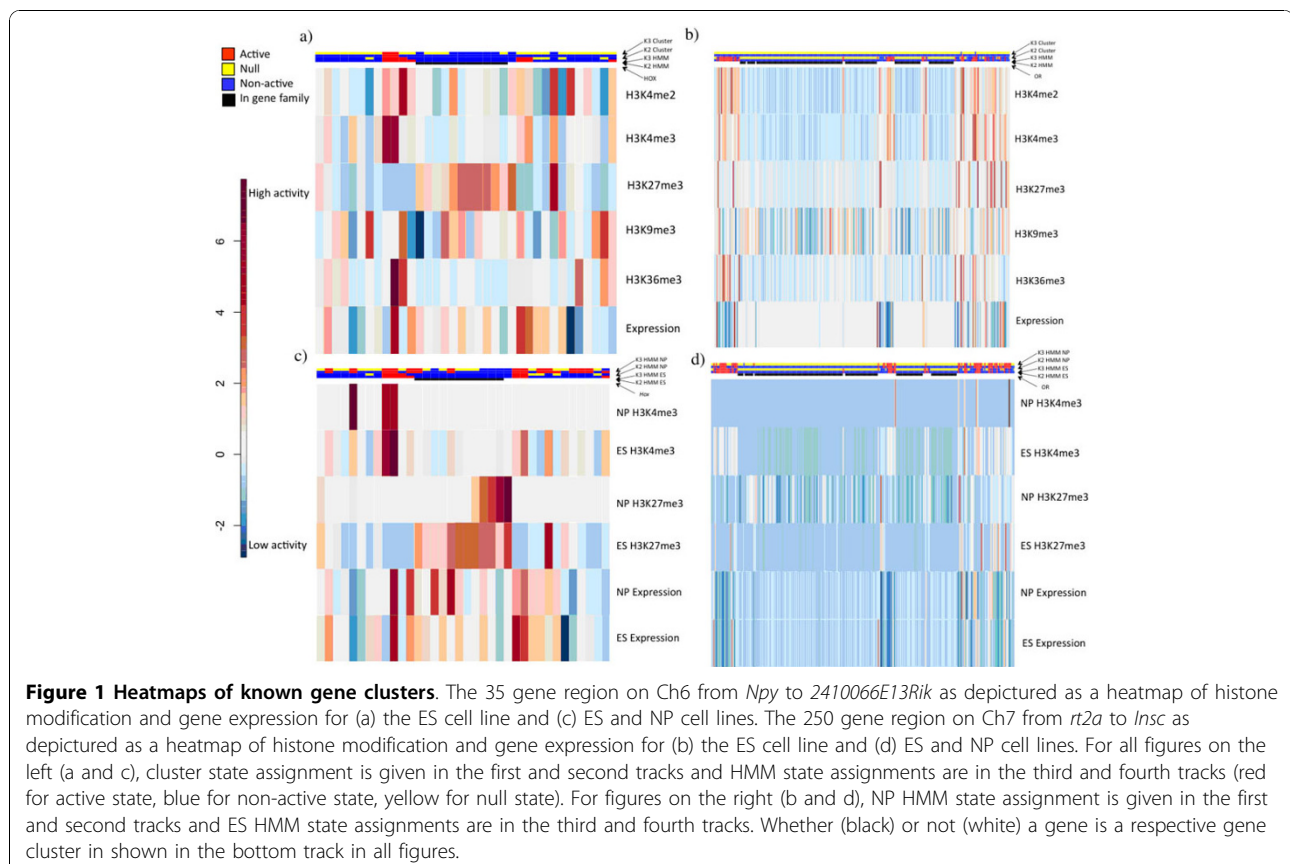
Jessica Larson^{1*}, Guocheng Yuan^{1,2}

From Sixth International Society for Computational Biology (ISCB) Student Council Symposium Boston, MA, USA. 9 July 2010

Motivation

Epigenetic regulations are important mechanisms for transcriptional control. There is evidence that neighbouring genes, although not always involved in the same

pathways, are still similarly regulated via various histone modifications. Currently, most studies are limited to local epigenetic patterns, whereas methods for analysing large-scale organizations are still lacking.



* Correspondence: jlarson@hsph.harvard.edu

¹Department of Biostatistics, Harvard University, Boston, MA 02115, USA

Full list of author information is available at the end of the article

Methods

We developed a computational approach to detect multi-gene domains with coherent epigenetic patterns. We applied this method to analyse a published ChIP-seq dataset for five different histone modification marks (H3K4me2, H3K4me3, H3K27me3, H3K9me3, H3K36me3) in mouse embryonic stem cells. We first obtained a 5-dimensional score for all known genes based on average modification activity in select regions. Then, with hidden Markov models and corresponding algorithms, we were able to determine the most probable domain status of each gene. We find that a three-state hidden Markov model can best describe the data, where the states correspond to active, inactive, and null domains.

Results

This model predicts 339 significantly large multi-gene domains, including known domains such as the olfactory receptor (OR) gene clusters, but also previously uncharacterized domains (Figure 1). We also noted less histone modification variability within each of our domains when compared to randomly selected boundaries. We further validated our predictions against gene expression and Gene Ontology data and found our domains were functionally relevant.

Conclusion

Our method provides a novel approach to analyse large-scale epigenetic patterns. As we continue to apply our method to other cell lines, we will provide important insight into the general structure, organization, and regulation of the mammalian genome.

Author details

¹Department of Biostatistics, Harvard University, Boston, MA 02115, USA.

²Department of Biostatistics and Computational Biology, Dana-Farber Cancer Institute, Boston, MA 02115, USA.

Published: 7 December 2010

doi:10.1186/1471-2105-11-S10-O5

Cite this article as: Larson and Yuan: A hidden Markov model for detecting multi-gene chromatin domains. *BMC Bioinformatics* 2010 11 (Suppl 10):O5.

Submit your next manuscript to BioMed Central and take full advantage of:

- Convenient online submission
- Thorough peer review
- No space constraints or color figure charges
- Immediate publication on acceptance
- Inclusion in PubMed, CAS, Scopus and Google Scholar
- Research which is freely available for redistribution

Submit your manuscript at
www.biomedcentral.com/submit

