

Poster presentation

Open Access

De novo identification of *in vivo* binding sites from ChIP-chip data

Victor Jin*¹, Alina Rabinovich², Henny O'Geen², Sushma Iyengar² and Peggy Farnham²

Address: ¹Bioinformatics Program and Department of Biology, University of Memphis, Memphis, TN 38152, USA and ²The Genome Center, University of California, Davis, Davis, CA 95617, USA

Email: Victor Jin* - vjin@memphis.edu

* Corresponding author

from UT-ORNL-KBRIN Bioinformatics Summit 2008
Cadiz, KY, USA. 28–30 March 2008

Published: 8 July 2008

BMC Bioinformatics 2008, 9(Suppl 7):P24 doi:10.1186/1471-2105-9-S7-P24

This abstract is available from: <http://www.biomedcentral.com/1471-2105/9/S7/P24>

© 2008 Jin et al; licensee BioMed Central Ltd.

Poster presentation

Advances in high-throughput technologies such as ChIP-chip and the completion of human genomic sequences allow analysis of the mechanisms of gene regulation on a systems level. In this study, we have developed a computational genomics approach (ChIPModules) and a motif discovery approach (ChIPMotifs) to mine the ChIP-chip data. The ChIPModules approach begins with experimentally determined binding sites and integrates positional weight matrices, a comparative genomics approach, and statistical learning methods to identify transcriptional regulatory modules. Using E2F1 ChIP-chip data performed on ENCODE regions in both HeLa and MCF7 cells, we have identified five regulatory modules for E2F1. One of modules was validated by using ChIP-chip with arrays containing ~14,000 human promoters. The ChIPMotifs approach incorporates a bootstrap re-sampling method to statistically infer the optimal cutoff threshold for a position weight matrix (PWM) of a motif identified from ChIP-chip data by *ab initio* motif discovery programs. Using OCT4 ChIP-chip data, we developed an *in vivo* OCT4 PWM. We then used this PWM and our ChIPModules to identify transcription factors co-localizing with OCT4 in a testicular germ cell tumor (Ntera2 cells).

V.X.J at the University of Memphis. As part of our analyses, we used ChIP-chip data collected as part of the ENCODE Project Consortium.

Acknowledgements

This work was supported in part by Public Health Service grant CA45250, HG003129, and DK067889 to P.J.F. and a bioinformatics start-up funding to