Poster A-6

Identification of conserved TAF1 promoters with functional roles in human Embryonic Stem cells

Authors:
Victor M. Ruotti (*WiCell Research Institute*)
Aaron Darling (*Dept. of Computer Science, University of Wisconsin-Madison*)
Ron Stewart (*WiCell Research Institute*)
Tae Hoon Kim (*. Ludwig Institute for Cancer Research, UCSD School of Medicine*)
Bing Ren (*. Ludwig Institute for Cancer Research, UCSD School of Medicine*)
James Thomson (*WiCell Research Institute*)

Short Abstract: We combine ChIP-chip hybridization data with a Mauve genome alignment of mammalian genomes to identify putative conserved TAF1 promoters and mammalian microrearrangements. In addition to previously reported TAF1 binding data, we utilized our own ChIP-on-chip array, which was designed to study promoters of human embryonic stem cells.

Long Abstract:

Genomic sequence data provides a framework for identification of biomarkers related to human embryonic stem (hES) cell state [6]. The existing annotation of the human genome, along with knowledge of human promoters, provides the basis for construction of the first “catalog” of promoters active in ES cells. Previous work has attempted to identify promoters with binding sites for the transcription factor TAF1, a subcomponent of the RNA polymerase II preinitiation complex (PIC) [5].

To better validate the location and function of putative TAF1 promoters we combined ChIP-chip hybridization data with a high-throughput comparative analysis of the human (NCBI 35), mouse (NCBI 33) and rat (RGSC 3.4) genomes. In addition to previously reported TAF1 binding data [6], we utilized our own ChIP-on-chip whole-genome tiling array, which was designed to study promoters of human embryonic stem cells (WiCell H1 cell line). We performed genome comparison using the Mauve genome alignment system (Darling et. al. 2004) to construct a mammalian homology map and alignment. The alignment forms a basis for further study into both conserved promoters and genomic rearrangements in mammalian genomes.

Although previous mammalian genomes alignments have been constructed [1-4], none have used a method sensitive to small rearrangements of genomic sequence. Mauve implements a novel method to identify genomic micro-rearrangements by anchoring alignments in conserved regions of non-repetitive sequence. Construction of the initial homology map consumed 12 hours on a Linux PC with two disks. Subsequent whole-genome alignment based on the initial homology map consumed another 12 hours on a 96-CPU Orion Multisystems deskside workstation. To evaluate the effect of micro-rearrangements of various sizes, we constructed alignments with increasing minimum LCB weights of 56, 100, 200, and 400. The alignments used for our analyses used a minimum LCB weight of 56 and resulted in 6351 LCBs in the initial homology map. Several thousand of these LCBs are less

than 1Kbp in length–truly micro-rearrangements.

This comparative map of the human, rat and mouse genomes, along with the information derived from our microarray study, allowed us to predict the location of putative functional promoters in hES cells and verify them as conserved across species. We further analyzed the possible role of micro-rearrangements of promoter sequences. We focused our analysis on TAF1 promoters, a subunit of the RNA polymerase II pre-initiation complex (PIC). Our comparative framework showed a high degree of conservation at the promoter level–75% of human TAF1 promoters were conserved in mouse and rat. Very few TAF1 promoters showed evidence for micro-rearrangement. Future work needs to be done to further annotate these conserved promoters. However, this analysis provides the framework to identify functional TAF1 promoters that may play an active role in hES cells.

References

[1] Strategies and Tools for Whole-Genome Alignments. Couronne, O., Poliakov, A., Bray, N., Ishkhanov, T., Ryaboy, D., Rubin, E., Pachter, L., Dubchak, I. 2002. Genome Res. 2003 Jan;13(1):73-80

[2] LAGAN and Multi-LAGAN: efficient tools for large-scale multiple alignment of genomic DNA. Brudno M, Do CB, Cooper GM, Kim MF, Davydov E, Green ED, Sidow A, Batzoglou S. Genome Research 2003 Apr;13(4):721-31

[3] Automated whole-genome multiple alignment of rat, mouse, and human. Brudno M, Poliakov A, Salamov A, Cooper GM, Sidow A, Rubin EM, Solovyev V, Batzoglou S, Dubchak I. Genome Res. 2004 Apr;14(4):685-92.

[4] Integrative genomics: in silico coupling of rat physiology and complex traits with mouse and human data. Twigger SN, Nie J, Ruotti V, Yu J, Chen D, Li D, Mathis J, Narayanasamy V, Gopinath GR, Pasko D, Shimoyama M, De La Cruz N, Bromberg S, Kwitek AE, Jacob HJ, Tonellato PJ. Genome Res. 2004 Apr;14(4):651-60

[5] A high-resolution map of active promoters in the human genome. Kim TH, Barrera LO, Zheng M, Qu C, Singer MA, Richmond TA, Wu Y, Green RD, Ren B.
Nature. 2005 Aug 11;436(7052):876-80.

[6] Genomic studies in stem cell systems. Zhao Y, Clark J, Ding S.
Curr Opin Mol Ther. 2005 Dec;7(6):543-50. Review.