1

In silico studies of the transcriptional regulation of the genes coding for the novel IL28A, IL28B, and IL29 protein family: A computational screening approach applicable on a genomic scale.

William Krivan,[1] Brian Fox,[2] Emily Cooper,[2] Teresa Gilbert,[2] Frank Grant,[2] Betty Haldeman,[2] Katherine Henderson,[2] Wayne Kindsvogel,[2] Kevin Klucher,[2] Gary McKnight,[2] Patrick O'Hara,[2] Scott Presnell,[2] Monica Tackett,[2] David Taft,[2] Paul Sheppard[2]

Keywords: interleukins, interferons, transcriptional regulation, phylogenetic footprinting

The novel IL28A, IL28B, and IL29 protein family consists of three non-allelic human proteins, and homologous mouse proteins, which are distantly related to interferons and IL-10 [1]. We use this protein family to illustrate an approach to the computational identification and characterization of putative transcriptional regulatory regions that consists of a combination of available and novel techniques (see [2] for a review) that can be applied on a genomic scale.

Insights into the regulatory mechanisms of the novel IL28A,B and IL29 protein family may be gained from comparisons of their potential regulatory regions with the regulatory regions of characterized cytokines such as IFN-alpha, beta, and gamma. In metazoans, however, it is in general not feasible to study co-regulation of paralogous genes by simply performing alignments of the upstream genomic sequences, an approach that has been successfully pursued for yeast and bacteria. Comparisons of potential regulatory regions must reveal subtle similarities such as individual transcription factor binding sites. However, the low binding specificity of transcription factors results in a high rate of false predictions in the computational analysis of genes from metazoan species. The number of predicted sites can be reduced by about one order of magnitude to a set more likely to have sequence-specific functions by means of phylogenetic footprinting, a conservation-based filter based on the biological observation that regulatory regions are often more highly conserved between species than other non-coding regions [3]. Another technique that can be used for the selection of presumably functional motifs is motivated by the observation that groups of transcription factors rather than single factors are required for biologically functional regulatory regions and is based on the hypothesis that statistical significance of clusters of sites is correlated with biological function [4, 5].

We illustrate the combined application of these techniques for the characterization of putative regulatory regions of IL28A,B and IL29. We also present results from genomic screens using a liver-specific model [6] and a model for immune-related gene function [7].

References

[1] Sheppard, P. et al. 2003. IL-28, IL-29 and their class II cytokine receptor IL-28R. *Nat. Immunol.* 4:63–68.

[2] Wasserman, W.W. and Krivan, W. 2003. In silico identification of metazoan transcriptional regulatory regions. *Naturwissenschaften* 90:156-166.

[1]ZymoGenetics, Inc. 1201 Eastlake Ave. East, Seattle, WA 98102, USA. E-mail: `krivan@zgi.com`
[2]ZymoGenetics, Inc. 1201 Eastlake Ave. East, Seattle, WA 98102, USA.

2

[3] Wasserman, W.W. et al. 2000. Human-mouse genome comparisons to locate regulatory sites. *Nat Genet.* 26:225-228.

[4] Wagner, A. 1999. Genes regulated cooperatively by one or more transcription factors and their identification in whole eukaryotic genomes. *Bioinformatics* 15:776–784.

[5] Frith, M.C., Li, M.C. and Weng, Z. 2003. Cluster-Buster: Finding dense clusters of motifs in DNA sequences. *Nucleic Acids Res.* 31:3666-3668.

[6] Krivan, W. and Wasserman, W.W. 2001. A predictive model for regulatory sequences directing liver-specific transcription. *Genome Res.* 11:1559-1566.

[7] Liu, R., McEachin, R.C. and States, D.J. 2003. Computationally identifying novel NF-κB-regulated immune genes in the human genome. *Genome Res.* 13:654-661.