A Whole-genome Analysis of Transcription Factor Binding Sites for Human and Mouse Orthologs

Caroline S Finnerty[1], Dr. James O McInerney[1]

Keywords: transcription factor binding sites, gene regulation, human and mouse orthologs.

1 Introduction.

We aim to test the hypothesis that dissimilarities in the control regions of genes are responsible for species-specific variation. We will do this by comparing human and mouse orthologs and analyse their upstream regions for presence or absence of various transcription factor binding sites (TFBS). We hope to see species-specific variation in TFBS that may be responsible at some level for species-specific differences. The biological significance of this work will be to determine if two genes which share a particular subset of transcription factors will have a similar function and if they will be expressed to the same level and conversely, if two genes have very divergent upstream sequences will their functions and expression levels also be dissimilar? The ultimate goal is to infer gene function from regulatory sequence.

2 Methods.

Our approach is to analyse, on a genome-wide scale the upstream regions of human genes with an emphasis on transcription factor binding sites. In order to take a conservative approach we have only used human genes, which have a corresponding mouse ortholog. This is commonly known as phylogenetic footprinting. Using existing databases such as TRANSFAC [2] to retrieve transcription factor binding site data we have recoded each transcription factor binding site with a different number so that each upstream region is identified by a different string of numbers. Using these newly recoded vectors, pairwise alignments using SWNumString.java (In-house software) have been carried out.

3 Discussion and Future Work.

These approaches should enable one to identify homologous upstream regions as well as those that are divergent which we can then analyse further.
We also hope to examine interspecies variation by performing multivariate analysis [3] on this dataset

[1]National University of Ireland, Maynooth, Co. Kildare, Ireland E-mail: caroline.s.finnerty@may.ie

4 References and bibliography.

[1] Dermitzakis, E.T. and Clarke, A.G. 2002. Evolution of transcription factor binding sites in mammalian gene regulatory regions: Conservation and turnover. *Molecular Biology and Evolution* 19(7):1114-1121

[2] Thioulouse, J., Chessel, D., Doledec, S., Olivier, J. M. 1997 ADE-4: A multivariate analysis and graphical display software *Statistics and Computing* 7(1)75-83

[3] Wingender et. al., 2000 TRANSFAC: an integrated system for gene expression regulation. *Nucleic Acids Research* 28(1):316-319