

304b Computational Analysis of Combinatorial Gene Regulation in the Liver

Joseph L. Vitolo, Ioannis (Yannis) P. Androulakis, and Charles M. Roth

Several approaches have been developed to derive greater meaning from gene expression profiles by incorporating knowledge of transcription factor binding sites on the promoters of target genes. For prokaryotic organisms such as *E. coli*, transcription of a particular gene is often controlled by one or a few operators, and databases of operons are growing. In contrast, mammalian gene regulation is particularly challenging due to the increased size of the “parts list”, the highly combinatorial nature of the regulation, and the complexities of post-transcriptional regulation mechanisms. We seek to obtain information on transcription factor regulatory sites on mammalian gene promoters and to use this for meaningful interpretation of gene expression changes in the liver under inflammatory conditions.

Beginning with the 8799 probes on the Affymetrix Rat Genome U34A Array, we were able to identify 2976 unique gene promoters in the rat genome using standard bioinformatics tools. Each promoter was analyzed in an 1100 bp window for the presence of transcription factor binding sites using position weight matrices, as implemented in the MatInspector program (Genomatix, Munich, FRG). This was performed for each of 374 matrices, falling into 142 families. The result is a dense matrix of promoters by matrices representing potential gene regulatory connections. We have characterized the statistics of transcription factor predictions across the gene promoter set as a function of stringency scores used in the identification of a transcription factor binding site as well as their locations relative to the transcriptional start site.

Our current work is focused on evaluating alternative methods for reducing the complexity of the predicted transcriptional network to a tractable size for interpretation of co-expressed groups of genes, both directly and in a mathematical framework of Network Component Analysis (NCA) (Liao et al., Proc. Natl. Acad. Sci., 2003). In addition to increasing stringency scores, we are assessing several potential methods for mining the TF prediction data set, including: restricting the spatial range of TFs considered; utilizing “literature modules” of demonstrated transcription factor combinations; significance testing of TF prediction frequencies in gene expression clusters vs. the whole genome; and the construction of sets of candidate transcriptional networks. The types of information that can be extracted from each of these approaches will be discussed using examples of gene expression changes in the liver after inflammatory insults and anti-inflammatory therapy. In particular, we identify transcription factors whose binding sites are highly represented among genes regulated in hepatic inflammatory responses and use these to reconstruct hepatic inflammatory regulatory networks.