

Learning where to look: Reducing the size of the haystack

Heather A. Amthauer, Costas Tsatsoulis, Electrical Engineering and Computer Science Department, University of Kansas, Lawrence, Kansas 66045.

2006 Annual Kansas City Area Life Sciences Research Day.

Looking for genetic networks can be equated to the proverbial task of looking for a needle in a haystack; or in the case of genetic networks, several needles in a haystack. Discovering and deciphering genetic networks is a complex task that can be made easier by providing researchers with a way to focus their searches for these networks within a genome. Recently, several studies have challenged the notion that gene location in higher eukaryotic organisms is random. These studies [Blumenthal, 1998; Boutanaev et al., 2002; Caron et al., 2001; Cohen et al., 2000; Lee & Sonnhammer, 2003; Lercher et al., 2002; Li et al., 2005; Mayor et al., 2004; Roy et al. 2002; Spellman & Rubin, 2002] suggest that there may be patterns in gene location. Currently, there is little work being dedicated to determining if there are true patterns within a genome, and if we can learn from these patterns. If there are patterns based on how genes cluster/group within a genome, we can utilize these patterns to assist in the discovery of genetic networks through inductive learning (process of learning by example). We are in the process of developing a sophisticated, computational tool that integrates different knowledge sources and learning algorithms to come up with reliable hypotheses about potential locations for genetic regulatory networks within a genome. The goal of this research is to learn where to look for genes that maybe coregulated with a given gene based on grouping patterns of known synexpression groups (synexpression group: set of genes that share complex spatial expression pattern and that function in the same process). Based on these grouping patterns (how large the group is, what other groups may be near it, etc), we could learn how to focus a search for genes that may be coregulated with a given gene.