

# **A METHODOLOGY FOR MOTIF DISCOVERY EMPLOYING ITERATED CLUSTER RE-ASSIGNMENT**

Osman Abul and Finn Drabl

Department of Cancer Research and Molecular Medicine, Norwegian University of Science and Technology, Trondheim, Norway

Geir Kjetil Sandve

Department of Computer and Information Science, Norwegian University of Science and Technology, Trondheim, Norway

Motif discovery is a crucial part of regulatory network identification, and therefore widely studied in the literature. Motif discovery programs search for statistically significant, well-conserved and over-represented patterns in given promoter sequences. When gene expression data is available, there are mainly three paradigms for motif discovery; *cluster-first*, *regression*, and *joint probabilistic*. The success of motif discovery depends highly on the homogeneity of input sequences, regardless of paradigm employed. In this work, we propose a methodology for getting homogenous subsets from input sequences for increased motif discovery performance. It is a unification of *cluster-first* and *regression* paradigms based on iterative cluster re-assignment. The experimental results show the effectiveness of the methodology.