Chapter 25
Comparison of Promoter Sequences Based on Inter Motif Distance

A. Meera
BMS College of Engineering, India

Lalitha Rangarajan
University of Mysore, India

ABSTRACT
Understanding how the regulation of gene networks is orchestrated is an important challenge for characterizing complex biological processes. The DNA sequences that comprise promoters do not provide much direct information about regulation. A substantial part of the regulation results from the interaction of transcription factors (TFs) with specific cis regulatory DNA sequences. These regulatory sequences are organized in a modular fashion, with each module (enhancer) containing one or more binding sites for a specific combination of TFs. In the present work, the authors have proposed to investigate the inter motif distance between the important motifs in the promoter sequences of citrate synthase of different mammals. The authors have used a new distance measure to compare the promoter sequences. Results reveal that there exists more similarity between organisms in the same chromosome.

1. INTRODUCTION
Common activities in bioinformatics/cognitive informatics include developing a unified analysis of pattern and organization of biological structures. Developing computational techniques that give insight into these areas is of utmost importance.

The hereditary information for organisms is carried in its genes. Genes are sequences of the polymer DNA which, for our purposes, can be viewed as strings over the alphabet \(\{A, C, G, T\}\), where each of the four characters corresponds to one of the nucleotide bases that makes up DNA. Individual genes are subsequences of the much larger strings of DNA that comprise the chromosomes of an organism. In addition to specifying
Comparison of Promoter Sequences Based on Inter Motif Distance

the structural information for proteins, genes must
be turned on and off at precisely the right time
and in the correct tissue in the developing and
mature organism. This process is termed as gene
regulation and is one of the central problems in
modern biology.

The first step in gene regulation is transcrip-
tion, where the information in a gene is amplified
by copying it into RNA, a polymer similar to
DNA. Short DNA sequences termed transcrip-
tion elements, typically of the order of 6-10 base
pairs in length, are recognized and bound by
sequence-specific binding proteins termed tran-
scription factors to form transcription complexes
through protein-protein as well as DNA-protein
interactions. Important transcription elements are
located immediately preceding the start of genes.
More surprisingly, transcription elements are also
found thousands of bases upstream, downstream
and even within the boundaries of a gene. The
transcriptional state of a gene (i.e., its time, tis-
 sue and rate of expression) is determined through
formation of a “transcription complex” composed
of multiple, interacting transcription factors bound
to their respective transcription elements. The
information needed to specify a transcription
factor binding site is not all local to an individual
transcription element, but requires interactions
with other binding sites through protein-protein
interactions to stabilize the complex.

In this paper, we propose to compare promoter
sequences by considering the important motifs that
are responsible for expression of that particular
gene. Some of the available tools that compare
promoter sequences are ConReal (Berezikov,
Guryev, & Cuppen, 2005) MUMmer (Kurtz, Phil-
lippy, Delcher, Smoot, Shumway, Antonescu, &
Salzberg, 2004). Pair wise comparison is possible
with these tools and also they do not provide a
similarity score. Another class of methods uses
prior knowledge of TFBSs to construct the align-
ments. While ConReal focuses on generating
an ordered chain of conserved TFBSs, thus not
aligning regions that do not contain them, Site-
blast is a BLAST (Michael, Dieterich, & Vingron,
2005)-like heuristic where the TFBS hits are used
as seeds. The method of Hallikas et al. (2006)
also falls in this category. Here, the sequence of
hit pairs is aligned using a scoring scheme that
considers clustering of sites, binding affinity and
conservation, though the underlying sequences
themselves are not aligned. Other approaches like
Monkey (Moses, Chiang, Pollard, Iyer, & Eisen,
2004) explicitly take into account evolutionary
properties of the TFBSs, but still perform the
alignment independent of the annotation step.

The focus of bioinformatics has begun to
extend from the identification of genes toward
understanding how the expression and regulation
of genes is orchestrated in a genomic level. Genes
expressed within the same biological context often
share promoter modules/frameworks (Fessele,
Maier, Zischek, Nelson, & Werner, 2002; Werner,

Understanding how the regulation of gene net-
works is orchestrated is an important challenge for
characterizing complex biological processes. Gene
transcription is regulated in part by nuclear factors
that recognize short DNA sequence motifs, called
transcription factor binding sites, in most cases
located upstream of the gene coding sequence in
promoter and enhancer regions. Genes expressed
in the same tissue under similar conditions often
share a common organization of at least some of
these regulatory binding elements. In this way
the organization of promoter motifs represents
a “footprint” of the transcriptional regulatory
mechanisms in a specific biologic context and
thus provides information about signal and tissue
specific control of expression. Promoters are the
central processors of transcriptional control, as the
regulatory information contributed by the other
elements must be integrated within the context of
a promoter in order to influence gene expression
(Werner, 1999). Understanding how networks
of promoters are organized provides insight into
when and how the expression of specific genes
is controlled.
Related Content

Ensemble of Neural Networks for Automated Cell Phenotype Image Classification
[www.igi-global.com/chapter/ensemble-neural-networks-automated-cell/56175?camid=4v1a](www.igi-global.com/chapter/ensemble-neural-networks-automated-cell/56175?camid=4v1a)

The Formal Design Model of an Automatic Teller Machine (ATM)
Yingxu Wang, Yanan Zhang, Philip C.-Y. Sheu, Xuhui Li and Hong Guo (2012). *Breakthroughs in Software Science and Computational Intelligence* (pp. 263-287).
[www.igi-global.com/chapter/formal-design-model-automatic-teller/64613?camid=4v1a](www.igi-global.com/chapter/formal-design-model-automatic-teller/64613?camid=4v1a)

Soft Subspace Clustering for Cancer Microarray Data Analysis: A Survey
[www.igi-global.com/chapter/soft-subspace-clustering-for-cancer-microarray-data-analysis/97056?camid=4v1a](www.igi-global.com/chapter/soft-subspace-clustering-for-cancer-microarray-data-analysis/97056?camid=4v1a)

Positive and Negative Innovations in Software Engineering
[www.igi-global.com/chapter/positive-negative-innovations-software-engineering/65133?camid=4v1a](www.igi-global.com/chapter/positive-negative-innovations-software-engineering/65133?camid=4v1a)