

Chapter 4

Promoter Structures Conserved between *Homo Sapiens*, *Mus Musculus* and *Drosophila Melanogaster*

Boris R. Jankovic

King Abdullah University of Science and Technology, Kingdom of Saudi Arabia

John A. C. Archer

King Abdullah University of Science and Technology, Kingdom of Saudi Arabia

Rajesh Chowdhary

Biomedical Informatics Research Center, USA

Ulf Schaefer

King Abdullah University of Science and Technology, Kingdom of Saudi Arabia

Vladimir B. Bajic

King Abdullah University of Science and Technology, Kingdom of Saudi Arabia

ABSTRACT

*Some of the key processes in living organisms remain essentially unchanged even in evolutionarily very distant species. Transcriptional regulation is one such fundamental process that is essential for cell survival. Transcriptional control exerts great part of its effects at the level of transcription initiation mediated through protein-DNA interactions mainly at promoters but also at other control regions. In this chapter, the authors identify conserved families of motifs of promoter regulatory structures between *Homo sapiens*, *Mus musculus* and *Drosophila melanogaster*. By a promoter regulatory structure they consider here a combination of motifs from identified motif families. Conservation of promoter structure across these vertebrate and invertebrate genomes suggests the presence of a fundamental promoter architecture and provides the basis for deeper understanding of the necessary components of the transcription regulation machinery. The authors reveal the existence of families of DNA sequence motifs that are shared across all three species in upstream promoter regions. They further analyze the relevance of our findings for better understanding of preserved regulatory mechanisms and associated biology insights.*

DOI: 10.4018/978-1-61350-435-2.ch004

INTRODUCTION

Some of the key processes in living organisms remain essentially unchanged even in evolutionary very distant species. Transcription regulation is one of such fundamental processes that are necessary for cell survival. Promoters are considered the key regulatory regions of gene transcription initiation and therefore would have the potential to be conserved in cases of highly essential functions. However, conservation of transcription regulation machinery across different species is not well understood (Wray et al, 2003).

Transcription controls location, timing and extent of gene expression in living organisms (Carey et al., 2009). Transcription regulation is significantly responsible for the species-specific patterns of development specificities (Wittcopp, 2010). In that context, study of conserved areas of promoter regions across species at varying evolutionary distance may help our understanding of regulatory mechanisms at play.

Availability of data related to genomic sequences as well as experimentally confirmed transcription start sites (TSS) for a number of species allows for a thorough investigation of shared elements in promoter regions between species. With that objective in mind, we attempt to identify shared families of motifs (short DNA sequences) between promoters of orthologous genes of two mammals, *Homo sapiens* and *Mus musculus*, and promoters of an invertebrate *Drosophila melanogaster* (fruit fly). Our approach is conservative: a) We first find motif families (MFs) shared between the three species; although this is a crucial first step it does not imply that there is a preserved regulatory mechanism across these three species; b) We determined the genes in human and mouse, as well as in fruit fly, whose promoters share a sufficient number (6 to 10) of MFs; we investigated in more detail functional properties of those groups of genes identified by having promoters that share eight MFs. c) We matched the consensus sequences of identified MFs with

the known binding sites of transcription factors, to find out the most common transcription factors involved. Our study is based on inferring preserved promoter structure via shared MFs across three species, and study of bio-chemical significance of the genes that share these structures across three species.

BACKGROUND

The three species in our study were chosen primarily because their genomic sequences and TSSs are well studied and documented in addition to their evolutionary relationship. We would expect that the number of shared MFs would decrease with evolutionary distance, but the nature and dynamics of this process is another interesting aspect of our study.

Studying of promoter motifs for *D. melanogaster* is reported in (Down et al., 2007). This study reveals through statistical analysis 87 novel motifs putatively involved in transcription factor binding. These motifs are determined using only fruit fly promoters and need not be conserved in evolutionary remote species. However, one can look at the shared MFs in promoter regions of several species. The basic idea is that presence of same MFs in promoter regions may suggest conserved regulatory functionality. Clearly, the number of shared MFs, together with the lengths of motifs implies statistical significance of any such commonality.

Organisms that are well studied as well as being evolutionary distant are particularly good candidates for such an analysis. Because it satisfies these requirements, the comparison between human and *Fly* promoters is particularly convenient. Because of these properties, this combination has been studied before (FitzGerald et al, 2006). That study evaluates conserved regions by detecting presence and frequencies of all possible octamers on close to 11,000 *Fly* promoters and two sets of human promoters, approximately 12,000–15,000 records. Their study finds that “*Fly* and human

11 more pages are available in the full version of this document, which may be purchased using the "Add to Cart" button on the publisher's webpage:

www.igi-global.com/chapter/promoter-structures-conserved-between-homo/60829

Related Content

Graph-Based Shape Analysis for MRI Classification

Seth Long and Lawrence B. Holder (2011). *International Journal of Knowledge Discovery in Bioinformatics* (pp. 19-33).

www.irma-international.org/article/graph-based-shape-analysis-mri/62299

Protein Secondary Structure Prediction Approaches: A Review With Focus on Deep Learning Methods

Fawaz H. H. Mahyoub and Rosni Abdullah (2024). *Research Anthology on Bioinformatics, Genomics, and Computational Biology* (pp. 1318-1341).

www.irma-international.org/chapter/protein-secondary-structure-prediction-approaches/342576

Proficient Normalised Fuzzy K-Means With Initial Centroids Methodology

Deepali Virmani, Nikita Jain, Ketan Parikh, Shefali Upadhyaya and Abhishek Srivastav (2018). *International Journal of Knowledge Discovery in Bioinformatics* (pp. 42-59).

www.irma-international.org/article/proficient-normalised-fuzzy-k-means-with-initial-centroids-methodology/202363

A Novel Deep Learning Method for Identification of Cancer Genes From Gene Expression Dataset

Pyingkodi Maran, Shanthi S., Thenmozhi K., Hemalatha D. and Nanthini K. (2024). *Research Anthology on Bioinformatics, Genomics, and Computational Biology* (pp. 62-77).

www.irma-international.org/chapter/novel-deep-learning-method-identification/342522

Information Needs and Assessment of Bioinformatics Students at the University of Swaziland: Librarian View

Satyabati Devi Sorokhaibam and Ntombikayise Nomsa Mathabela (2017). *Library and Information Services for Bioinformatics Education and Research* (pp. 188-198).

www.irma-international.org/chapter/information-needs-and-assessment-of-bioinformatics-students-at-the-university-of-swaziland/176144