

In Silico approach to identify transcription factor binding sites and Cis-regulatory elements in tubulin gene promoter

Hira Mubeen, Shahid Raza

University of South Asia, Lahore, Pakistan

Abstract- Transcription factor (TF) binding sites or motifs (TFBMs) are functional *cis*-regulatory DNA sequences that play an essential role in gene transcriptional regulation. TFs play important roles in cellular physiology, developmental processes and responses to environmental stimuli. The importance of promoter analysis relies on how abundantly TFBSs and TFs are integrated into reference databases. The primary components of promoter are *cis* acting regulatory regions. Various tools and software's are being used to analyze and predict regulatory motif within promoter region of a gene. PlantPAN serves as a systematic platform for plant promoter analysis and reconstructs transcriptional regulatory networks. In the present study, we identified various motif including: light responsive elements, *cis* regulatory elements, core promoter elements and transcription factor binding motifs on the enhancer sequences of tubulin gene participating in variety of functions occurring in cellular processes.

Index Terms- gene regulations, motif, core promoter, TFB's.

I. INTRODUCTION

Gene expression regulation is essential for all cellular processes with proper transcriptional control. Gene expression is the process in which gene is decoded to form an mRNA sequence to produce the particular protein sequence. For this, a molecule called a transcription factor is needed. Transcription factors are actively involved in gene expression regulation and play an important role in the life cycle of all living

organisms including animals and plants. They can act as sequence specific DNA binding proteins which will bind with promoter region of gene to make it express. However, it is difficult to discover many *cis*-regulatory elements with existing software's because many transcription factors are still unknown. [1,2]

Studies on transcription factors (TFs) and *cis*-acting elements in promoters have attained much attention due to their important role in gene expression and its regulation. Therefore, identifying all functional binding sites within a known promoter is difficult and also the existence of some additional binding sites should be assumed. TFs recognize specific *cis*-regulatory sequences in promoter region to regulate target genes. In plants, TFs play an important role in many cellular processes and developmental processes. Also, their role in transcription regulation can never be ignored. Prediction of genes and their regulation on the basis of TF binding sites in the promoter region is considered as a main concern in system biology.

For prediction of genes, several useful web-based resources were established. For instance, TRANSFAC is the commercial database manually collecting experimentally verified TFs, transcription factor binding sites (TFBSs) and matrix-based target prediction profiles [3,4]. Two other well-known TFBS repositories, PLACE and JASPAR, were used to facilitate the identification of TFBSs in input sequences given by users [5,6]. However, the information on TFBSs is only provided rather than their corresponding TFs.

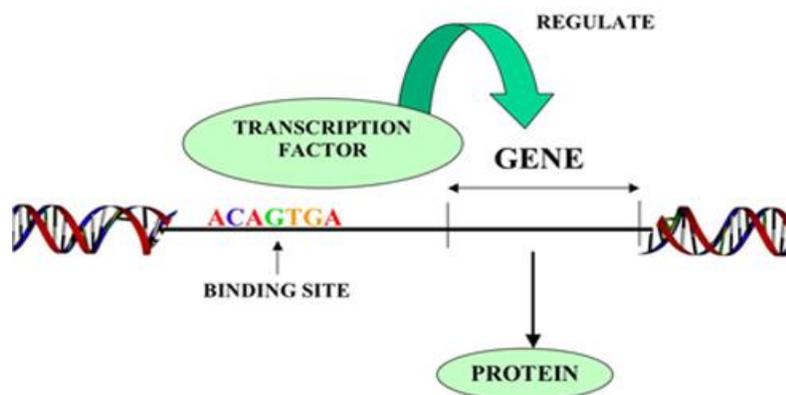


Fig 1: A transcription factor molecule binds to DNA at its binding site and regulates the production of a protein from its gene.

Tubulin

Tubulin is a globular protein that is considered as a building block of microtubules, which are a major element of the cytoskeleton. Tubulin heterodimers are composed of two major classes including alpha and beta tubulin. According to previous studies, tubulin genes of a wide variety of organisms have been isolated and characterized since the cloning of chicken and tubulin cDNA during the 1980s. Identification of functional transcripts has helped to classify tubulin into further classes and many isoforms. In addition, the complexity of the tubulin population is dependent on transcriptional regulation and post-translational modifications. These post-translational modifications allowed subdividing isoforms into several isoforms. [7,8,9,10]

Cis-Regulatory Module

The TFs bind in the regulatory regions of genes mainly enhancers organized in the form of modules like cis-regulatory modules (CRM), to sequences called as transcription factor binding sites (TFBS). CRM share regulatory sequences located few kilo bases away from gene of interest and bind to specific TFs at specific developmental stage to result in specific cell specification. [11] Overall, gene expression is regulated by the combination of all CRMs acting on genes throughout the organism's life. Studies shown, there exists as many as 10-fold more CRMs than genes. [12] Interaction between the TFs and CRMs form a development transcriptional regulatory network, which helps to carry out the activities for specification and differentiation programs of various cell types for development of organism.

PlantPAN

PlantPAN is software used as an informative resource for detecting transcription factor binding sites, corresponding transcription factors, and other important regulatory elements in a promoter or a set of plant promoters. Additionally, TFBSs, CpG islands, and tandem repeats in the conserve regions between similar gene promoters are also identified. The new PlantPAN (version 2.0) contains 16 960 TFs and 1143 TF binding site matrices among 76 plant species. It can help to update the annotation information, adding experimentally verified TF matrices, and making improvements in the visualization of transcriptional regulatory networks.

In the present study, we have identified putative transcription factor binding motifs on the enhancer sequences of tubulin gene participating in variety of functions occurring in several cellular and developmental processes. These mapped sequences are of great importance in finding out other regions within the genome having the similar or identical set of motifs. The importance of having these motifs mapped will help to know about which transcription factor will bind to particular promoter sequence. This connection between the promoter sequences and transcription factors regulate the gene expression.

II. MATERIAL AND METHODS

TFB Prediction

Transcription factor (TF) binding site prediction remains a challenge in gene regulatory research due to degeneracy and potential variability in binding sites in the genome.

PlantPAN

Plant promoter analysis navigator, for identifying combinatorial cis-regulatory elements with distance constraint in plant gene groups. <http://plantpan2.itps.ncku.edu.tw/>.

PLACE

Database of motifs found in plant cis-acting regulatory DNA elements, all from previously published reports. It covers vascular plants only. <http://www.dna.affrc.go.jp/htdocs/PLACE/>.

Motif Analysis

Motif analysis involves identification of common patterns from promoter region without knowing their binding sites.

PlantCARE

PlantCARE, a database of plant cis-acting regulatory elements and a portal to tools for *in silico* analysis of promoter sequences. <http://bioinformatics.psb.ugent.be/webtools/plantcare/html/>.

III. RESULTS

Analysis of Transcription Factor Binding Sites (TFB's)

PlantPAN software (www.plantpan.mbc.nctu.edu.tw/) identifies the transcription factors that are key regulators of gene expression. The putative transcription factor binding sites (TFBs) in Beta Tubulin (BT) gene promoters are illustrated below in table 1.

Table 1: Transcription factor binding sites

Transcription Factor	Species	Sequence	Site
Athb-1	Arabidopsis	gtgcATAATaagga	117
ANT	Arabidopsis	taacgCCCGatt	221
AG	Arabidopsis	aactgaTTGGtga	887
AGL3	Arabidopsis	acgttaTATGGtct	881
Athb-1	Arabidopsis	tTcATAATcaag	174

Motif Analysis

Light Responsive Elements

Analysis of beta tubulin gene promoter reveals various light responsive, core promoter elements and cis-regulatory elements. Some of them are given below.

Table 2: Shows light responsive elements

Motif	Species	Sequence
GAG-motif	<i>Arabidopsis thaliana</i>	AGAGAGT
I-box	<i>Solanum tuberosum</i>	TATTATCTAGA

Table 3: Shows core promoter elements

Motif	Species	Sequence
TATA-box	<i>Arabidopsis thaliana</i>	TATAAA
TATA-box	<i>Brassica napus</i>	ATATAT

Table 4: Shows cis-regulatory elements

Motif	Species	Sequence
5UTR Py-rich stretch	<i>Lycopersicon esculentum</i>	TTTCTTCTCT
ACE	<i>Petroselinum crispum</i>	CTAACGTATT
TC-rich repeats	<i>Nicotianatabacum</i>	ATTTTCTTCA

IV. DISCUSSION

Many *cis*-regulatory motifs were detected in the beta tubulin gene promoter. TC-rich repeats are important *cis*-acting element involved in defense response. Beta tubulin gene promoter was found to be enriched with conserved Transcription binding site motifs AGL3, AG, Athb-1, ANT. These motifs are located in sense (+) and (-) strands of tubulin promoter sequence. Major transcription binding sites found are AGL3, AG, Athb-1, ANT. The transcription factor AGL3 is expressed in all above-ground vegetative organs. It was observed that AGL3 may be involved the transcriptional regulation of genes. The prediction of regulatory motifs, TFBS form an essential link in comparative genomics. These sequences are evolutionary conserved and eventually we can find out the orthologous of these genes in higher and complex organism which help in understanding molecular mechanisms. But some of the hurdles to predictions mainly includes their location i.e, these modules are located far away from the genes they regulate. Next, the presence of multiple transcription factor binding sites for various TFs leads to combinatorial control of gene regulation, thus making it difficult to associate with one gene.^[13,14]

V. CONCLUSION

A large proportion of gene regulations happen at the transcriptional level through the binding of transcription factors to short regulatory DNA sequences in the upstream regions of TF target genes i.e, promoter. The problem of computationally identifying TF binding sites and motifs (TFBMs) remains a challenge. Hence it is suggested to develop and maintain more advances tools to analyze and identify the *cis* regulatory elements dispersed over promoter regions of genes.

REFERENCES

- [1] H. Leung, F. Chin, S.M. Yiu, R. Rosenfeld and W.W. Tsang, *JCB*, 12(6), 686-701 (2005).
- [2] M. Li, B. Ma, and L. Wang, *Journal of Computer and System Sciences*, 65, 73-96 (2002).
- [3] Matys, V., Kel-Margoulis, O.V., Fricke, E., Liebich, I., Land, S., Barre-Dirrie, A., Reuter, I., Chekmenev, D., Krull, M., Hornischer, K. et al. (2006) TRANSFAC and its module TRANSCOMP: transcriptional gene regulation in eukaryotes. *Nucleic Acids Res.*, 34, D108–D110.
- [4] Wingender, E. (2008) The TRANSFAC project as an example of framework technology that supports the analysis of genomic regulation. *Bioinformatics*, 9, 326–333.
- [5] Mathelier, A., Zhao, X., Zhang, A.W., Parcy, F., Worsley-Hunt, R., Arenillas, D.J., Buchman, S., Chen, C.Y., Chou, A., Ienasescu, H. et al. (2014) JASPAR 2014: an extensively expanded and updated open-access database of transcription factor binding profiles. *Nucleic Acids Res.*, 42, D142–D147.
- [6] Higo, K., Ugawa, Y., Iwamoto, M. and Korenaga, T. (1999) Plant *cis*-acting regulatory DNA elements (PLACE) database: 1999. *Nucleic Acids Res.*, 27, 297–300.
- [7] Cleveland L, MacDonald C, Rutter K (1980). Number and evolutionary conservation of α - and beta-tubulin and cytoplasmic beta- and gamma-actin genes using specific cloned cDNA probes. *Cell*. 20: 95-105.
- [8] Cleveland L, Sherline K (1981). Unpolymerized tubulin modulates the level of tubulin mRNAs. *Cell*. 25: 537-546.
- [9] Luduena (1998). Multiple forms of tubulin: different gene products and covalent modifications. *Int. Rev. Cytol.* 178: 207-275.
- [10] MacRae (1997). Tubulin post-translational modifications--enzymes and their mechanisms of action. *Eur. J. Biochem.* 244: 265-278.
- [11] Girardot C. Deciphering enhancer activity in *Drosophila* based on transcription factor occupancy and chromatin state chromatin state characterization (Doctoral dissertation, Université Pierre et Marie Curie-Paris VI). 2012.
- [12] Bryantsev AL, Cripps RM. Cardiac gene regulatory networks in *Drosophila*. See comment in PubMed Commons below *BiochimBiophysActa*. 2009; 1789: 343-353.
- [13] Berman BP, Pfeiffer BD, Laverty TR, Salzberg SL, Rubin GM, Eisen MB, et al. Computational identification of developmental enhancers: conservation and function of transcription factor binding-site clusters in *Drosophila melanogaster* and *Drosophila pseudoobscura*. *Genome Biol*. 2004; 5: R61.
- [14] Berman BP, Nibu Y, Pfeiffer BD, Tomancak P, Celniker SE, Levine M, et al. Exploiting transcription factor binding site clustering to identify *cis*-regulatory modules involved in pattern formation in the *Drosophila* genome. See comment in PubMed Commons below *Proc Natl AcadSci U S A*. 2002; 99: 757-762.

AUTHORS

First Author – Hira Mubeen, University of South Asia, Lahore, Pakistan, email: hira_sh@hotmail.com

Second Author – Shahid Raza, University of South Asia, Lahore, Pakistan