Transcription factor information system: A tool for detection of transcription factor binding sites

Transcription Factors are proteins those bind to DNA at specific sites (Transcription Factor Binding Sites or TFBSs) and control the gene expression. Hence identifying these TFBSs in DNA is important to understand the mechanism that regulates gene expression. For efficient understanding of these mechanisms there is a need to develop new computational tools for detecting these TFBSs, having more efficiency and advanced options so that detailed hence better results can be obtained.

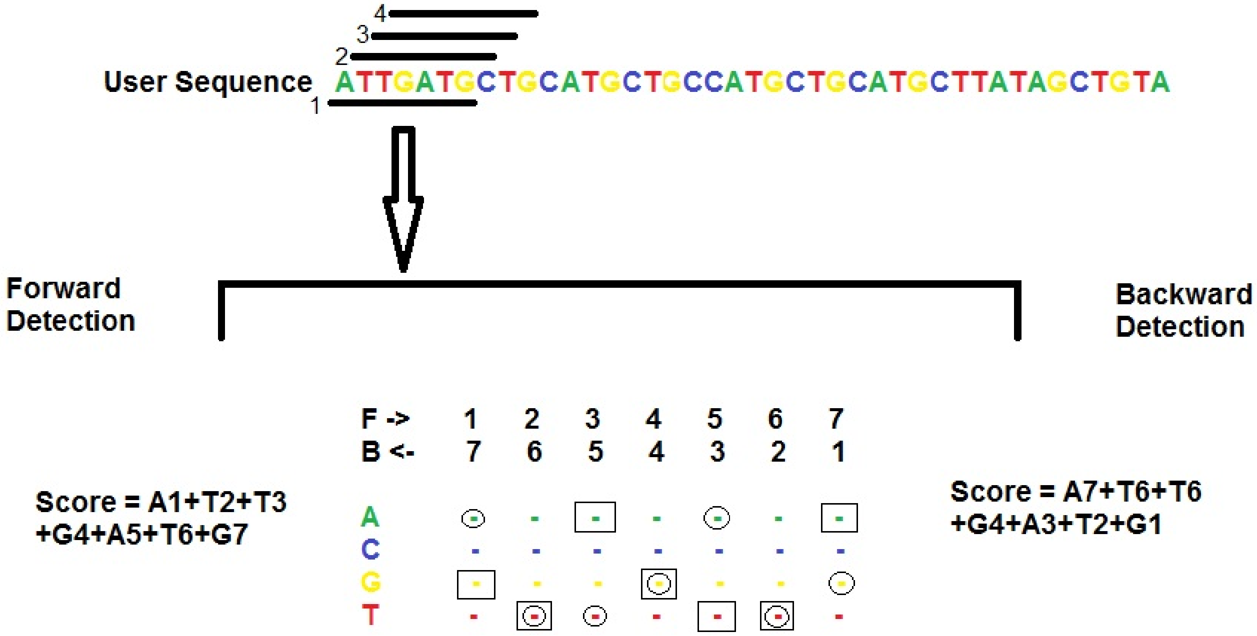

Fig. 1.

These tools must be developed as web applications so that user can easily access the capabilities online. These tools must also have user friendly interface so that a person with biology background can understand the functioning and result output easily. For making these identifications easier via computational methods we have developed a new online tool called Transcription Factor Information System (TFIS) which is based on position weight matrix for detecting TFBSs in a given DNA sequence. TFIS uses position frequency matrix library derived from JASPAR database for obtaining position weight matrix via position probability matrix. This tool provides better chance to scan a gene sequence for a variety of TFs. Although a lot of similar tools are available but TFIS provides certain unique and advanced features developed according to the requirement of the user. TFIS uses latest libraries of PFM and a global pseudo count of 0.8. It also employs a new method for threshold score determination based on percent scoring where cut-off is calculated with respect to the highest score corresponding to the best match for a particular TF.

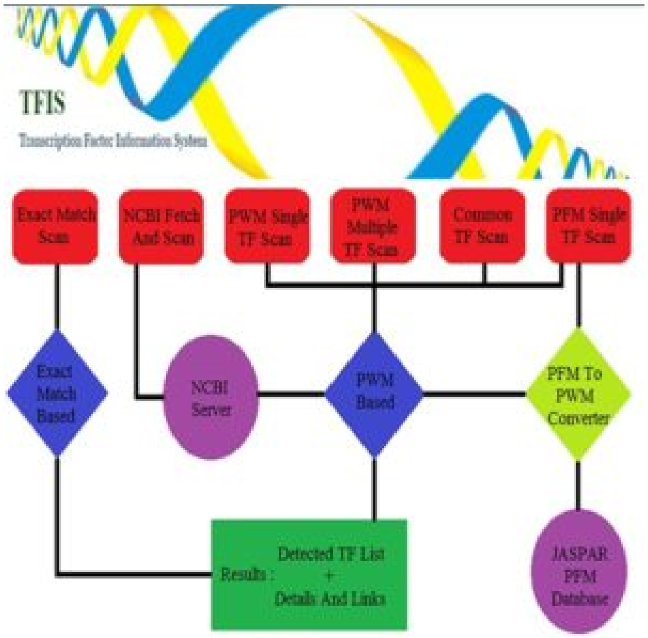

Fig. 2.

Advanced features include common TF scanning from a batch of sequences, specific TF scanning, NCBI database connectivity, manual scanning using user defined sequences and converting them to PWM etc. Results obtained using our tool are also more user friendly and informative for example, TFIS tool specifies position number in addition to highlighting the site in sequence and also links TFs to other important websites and databases like UNIPROT and NCBI. TFIS can detect the presence of potential TFBSs in both the direction at the same time. This feature increases its efficiency. And the results for this dual detection are presented in different colours specific to the orientation of the binding site. TFIS is equipped with advanced features such as direct information fetching from NCBI database, detecting binding site for common TF in a batch of gene sequences and TFBS detection after generating PWM from raw sequences in addition to general detection methods. This tool supports the use of JAVA language as an advanced bioinformatics programming language. This tool can be publicly accessed at http://www.tfis.in. This tool is surely an advanced tool as it is able to provide solutions which are not present in any other tool for example, TFSEARCH tool has a limit of up to 9,999 bases for gene sequence input while TFIS is free of such limitation. Results obtained by the TFIS are more detailed and specific to the detected TFs as integration of more informative links from various related web servers are added in the result pages like Gene Ontology, Pazar database and Transcription Factor encyclopaedia in addition of NCBI and UniProt.

Priyanka Narad
Amity Institute of Biotechnology, Amity University Uttar Pradesh, U.P., India

Publication

Transcription Factor Information System (TFIS): A Tool for Detection of Transcription Factor Binding Sites.
Narad P, Kumar A, Chakraborty A, Patni P, Sengupta A, Wadhwa G, Upadhyaya KC
Interdiscip Sci. 2016 Apr 6