

Introduction

D-Light on Promoters is a simple yet powerful client-server based system for calculating, querying and viewing transcription factor binding site (TFBS) annotation data on promoter sequences. The software provides scripts to set up an initial annotation dataset based on promoter sequences derived from the UCSC genome database[1] and position frequency matrices (PFMs) obtained from the JASPAR database[2]. Users can create personal accounts to add their own PFMs and/or promoter sequences for on-the-fly annotation.

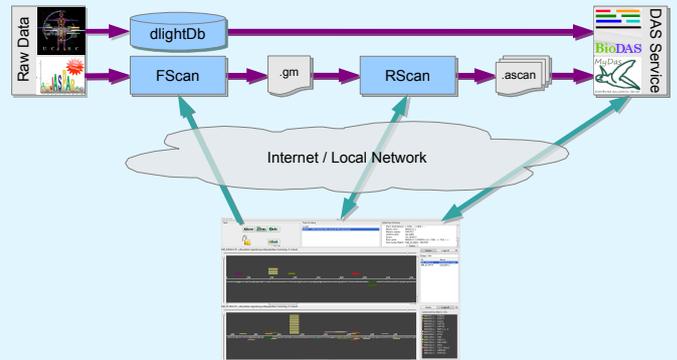
Background

The transcription of genes is controlled by the interplay of **transcription factors** (TFs) attached to their TFBSs on cis-regulatory elements. A **PFM** represents the binding site characteristic of a certain TF. PFMs are based on aligned binding site sequences derived by biological experiments. The frequency of occurrence for each base A,C,T or G is counted in each alignment column. The PFMs are subsequently used to predict potential binding sites for the corresponding TFs.

aaGTAACAATgc
 agGTAACAATgc
 gtGTAACAATgc
 atGTAACAATgc
 tGTAACAATgc

A	[1 0 19 20 18 1 20 7]
C	[1 0 1 0 1 18 0 2]
G	[17 0 0 0 1 0 0 3]
T	[1 20 0 0 0 1 0 8]

Implementation



D-Light on Promoters consists of five main modules:

- ▶ **dlightDb**: MySQL database with annotations for PFMs and promoter sequences
- ▶ **FScan**: precalculates binding sites
- ▶ **RScan**: searches in FScan results
- ▶ **DAS server**: embeds scan results and annotations into a DAS[3] conform XML format (based on MyDas[4])
- ▶ **Client**: result visualization, data upload and annotation scans

User Interface

The **D-Light Client** is available as Java Applet or standalone Java application and provides views to (i) scan the annotation set, (ii) select scan results, (iii) search in the dlightDb database, and (iv) visualize the results. The visualization is based on the GenoViz SDK[5]. Results can be locally stored as text file or as image.

The screenshots illustrate the user interface workflow:

- Select Ascan Result for Viewing**: A window for selecting scan results from a list of files.
- Search in Datasets**: A search interface with criteria like 'FOX*' and checkboxes for Matrix ACC, Gene ACC, Result Files, Matrix Keywords, Gene Keywords, and Setup.
- Scan Annotation**: A configuration window for matrix selection (e.g., MA0031), gene selection, genome (human/rat), and output options.
- Upload Matrix and Annotate**: A window for uploading user-defined matrices and sequences.
- Main Visualization**: A central window showing search results, task history, and genomic tracks with gene annotations and matrix overlays.

D-Light provides the opportunity to upload user-specified PFMs or DNA-sequences to extend the existing dataset. Unlike the initial dataset, all user-defined data are private and hidden from the other users.

Conclusion

D-Light on Promoters is an easy to use client-server system for managing and querying TFBS annotation data. The client is written in Java to assure platform independency. The server only requires a Linux system. The built-in annotation method applies a simple scoring scheme. The modular setup allows the replacement of FScan by other scoring tools. D-Light is freely available and licensed under GNU General Public License, version 3.

References

- [1] The UCSC Genome Browser database: update 2010. Rhead B *et al.* Nucleic Acids Res. 2010. <http://genome.ucsc.edu/>
- [2] JASPAR 2010: the greatly expanded open-access database of transcription factor binding profiles. Portales-Casamar E *et al.* Nucleic Acids Res. 2010. <http://jaspar.cgb.ki.se/>
- [3] The distributed annotation system. Dowell RD *et al.* BMC Bioinformatics 2001. http://www.biodas.org/wiki/Main_Page
- [4] MyDas <http://code.google.com/p/mydas/>
- [5] GenoViz Software Development Kit: Java tool kit for building genomics visualization applications. Helt GA *et al.* BMC Bioinformatics 2009. <http://genoviz.sourceforge.net/>