WEB-BASED TOOL FOR ANALYSIS OF SNPS IN NONCODING DNA

Lobanova M., Konovalova T., Cheremushkin E.
Institute of Cytology and Genetics SB RAS, Novosibirsk

Lobanova M.
Konovalova T.
Cheremushkin E.
Institute of Cytology and Genetics SB RAS

Abstract

We have created a tool for the SNP analysis in regulatory sequences. Those SNP which occurred in the potential sites of the Transcriptional Factor (TF) binding were of special interest. This application could be used as standalone program, as well as web-based tool. This program supports the following formats: SNPs – Fasta rsSNPs, promoters – Ensembl, weight matrices of TFs - TRANSFAC®. This tool allows the fast and convenient creation of programs for solving the related problems in the areas of molecular biology and diagnostics.

Project description

Introduction: The presence of SNP in the regulatory sequence may result in different levels of gene expression for distinct haplotypes. This influence may be based on difference in the TF binding with it's site. Provided with the information of SNPs found in potential TF-binding sites, one may estimate the influence exercised by these TFs on gene expression. Further, one may predict the level of expression of various genes in the samples with known SNPs.

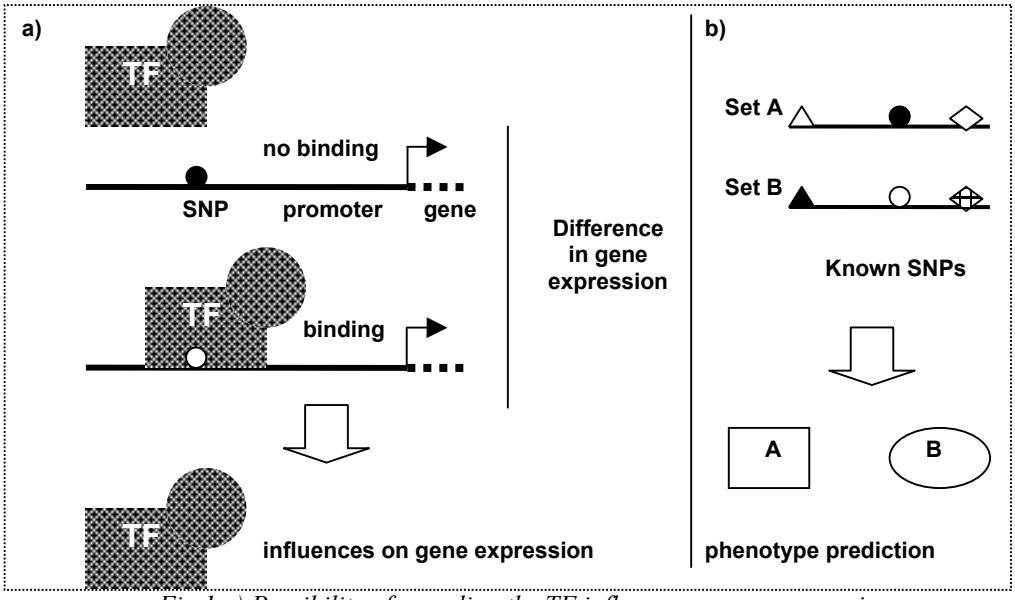

Fig.1 a) Possibility of revealing the TF influence on gene expression

b) further development of this concept – phenotype prediction based on SNP knowledge

Project description*:* SNP research tool is made on the PHP language. It's basic structure includes the following classes:

Matrix – the weight matrix revealing the TF affinity to the DNA sequence. Matrices should be in TRANSFAC® format.

Promoter – sequence with typical length of 1000-10000 nucleotides taken from the regulatory part of gene. The input file should be in the Ensembl format.

SNP – contains information about different alleles, SNP position

Site – are determined basing on the matrices

Supplementary classes.

Web interface: Web interface on this program may be found at: http://compel.bionet.nsc.ru/snpresearch/web/request_form.php

It's organized as follows: on the first stage there are alternatives of information input:

Example sequence; Paste SNP in Fasta Rs format; Load Rs SNP from file; Paste plain sequence with corresponding polymorphism position and alleles.

On the next stage the profile is chosen which represents set of TRANSFAC® weight

matrices for Transcription Factors with pre-selected matrix and core cutoffs. Then the result is displayed, which includes sets of sites that were found for both SNP alleles with their difference emphasized.

Further development

We plan to integrate the program performing the SNP research with other programs of our team. Also the processing of experimental data is to be begun.

Literature

1. *Kel A.E., Gossling E., Reuter I., Cheremushkin E., Kel-Margoulis O.V., Wingender E.* MATCH: A tool for searching transcription factor binding sites in DNA sequences.
2. *Djordjevic M., Sengupta A.M., Shraiman B.I.* A biophysical approach to transcription factor binding site discovery
3. *Wingender E., Chen X., Fricke E., Geffers R., Hehl R., Liebich I., Krull M., Matys V., Michael H., Ohnhäuser R., Prüß M., Schacherer F., Thiele S. and Urbach S.* "The TRANSFAC® system on gene expression regulation." Nucleic Acids Res. 29, 281-283 (2001)