The TRANSFAC® Database

Dating back as early as 1988, when the first data collection of transcription factors (TFs) and their binding sites was published [Wingender, Nucleic Acids Res. 16:1879-1902, 1988], TRANSFAC® has been the first and is the most comprehensive database on eukaryotic transcription regulation. It has been merged with TRANSCompel (a database on composite elements) and TRANSpro (a comprehensive collection of promoters of human and eight other genomes).

TRANSFAC ® is now also available under the geneXplain platform, providing the most comprehensive collection of TF DNA-binding profiles available for the state-of-the-art sequence analysis implemented in the platform.

GeneXplain offers a one-stop shopping solution for the platform together with the TRANSFAC® database as well as stand-alone solutions.

Applications

The most popular application of TRANSFAC contents is the prediction of potential transcription factor binding sites (TFBSs). Its contents can also be used to train own pattern finding algorithms, or to mine the wealth of information about transcription factors.

Further reading

Wingender, E. (2008) The TRANSFAC project as an example of framework technology that supports the analysis of genomic regulation. *Brief. Bioinform.* 9:326-332.

TRANSFAC®

About geneXplain

GeneXplain's mission is to provide a comprehensive platform for bioinformatic, systems biological and cheminformatic tools. The raison d'être of this platform is to assist translational research in the life sciences, mainly in the context of personalized medicine and pharmacogenomics. We intend to make our expertise available to academic and commercial partners in collaborative research projects.

To achieve this, geneXplain offers:

- The geneXplain platform providing a large number of bioinformatic and systems biological data analysis workflows. Unique is geneXplain's Upstream Analysis for causal interpretation of expression data.
- TRANSPATH®, one of the largest pathway/network databases presently available, particularly well suited for geneXplain's proprietary *Upstream Analysis*.
- HumanPSD, a rich information resource connecting pathways with targets, drugs and clinical trials.
- PASS and PharmaExpert for predicting biological activities of compounds qualitatively
- GUSAR for QSAR model building and quantitative activity prediction

geneXplain GmbH
Am Exer 10b
D-38302 Wolfenbüttel, Germany

info@genexplain.com
www.genexplain.com
www.facebook.com/genexplain
www.linkedin.com/company/genexplain
www.gplus.to/geneXplain

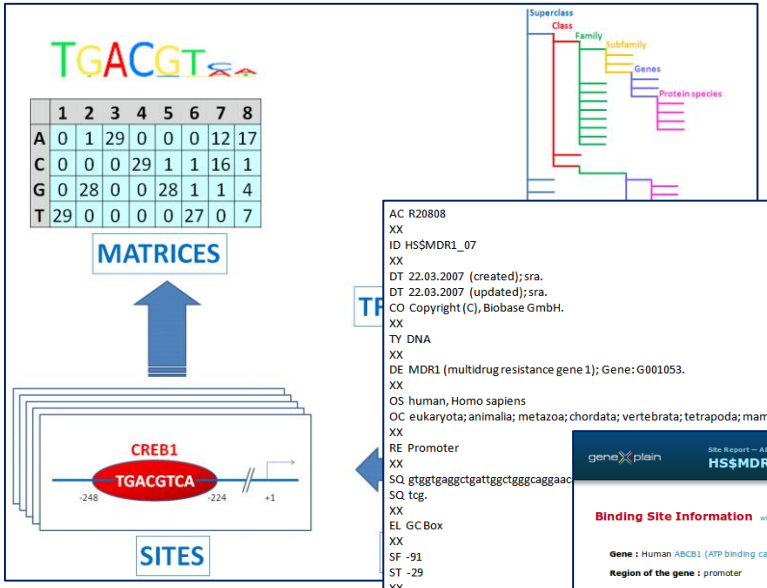
Directors: E. Wingender, A. Kel • Commercial register: HRB 202564, Amtsger. Braunschweig • VAT No.: DE271983408

TRANSFAC

The database of transcription factors, their genomic binding sites, and DNA-binding profiles

geneXplain

TRANSFAC®: the database on eukaryotic transcription

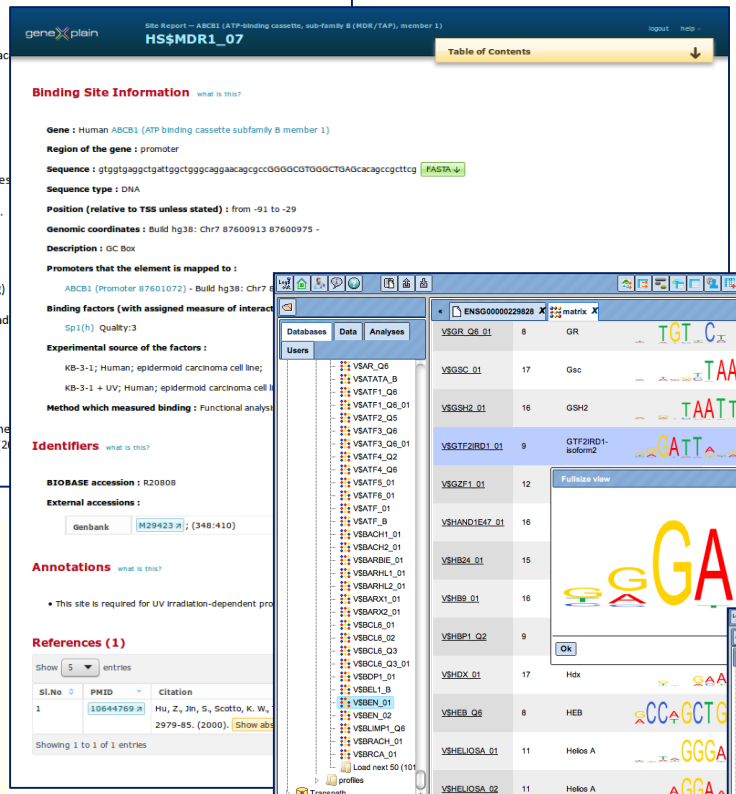


Key features

- 68,000+ transcription factor binding site reports containing details from the primary literature for more than 300 species, with a focus on human, mouse, rat, yeast, and plants
- 24,500+ transcription factor and miRNA reports, a subset of which provide GO functional assignments, disease associations and expression pattern assignments
- 30,300,000+ ChIP fragment reports that include the best scoring site prediction for the respective factor as well as downloadable sequences and gene lists
- 323,000+ promoter reports including ChIP-chip/Seq based histone modification locations, transcription start sites, and single nucleotide polymorphisms (SNPs)
- A pathway visualization tool for building custom regulatory networks out of experimentally demonstrated factor-DNA and factor-factor interactions

Site & promoter analysis

Using the rich library of 6500+ positional weight matrices of the TRANSFAC® database, DNA sequences can be scanned for potential transcription factor binding sites. One option for this is the proven tool Match™, which comes along with a standard TRANSFAC ® license, or to use one of the new sophisticated tools that are additionally provided by the geneXplain platform.



The structure

The core of TRANSFAC® comprises contents of two domains: Eukaryotic transcription factors (TFs) and TF binding sites (TFBSs).

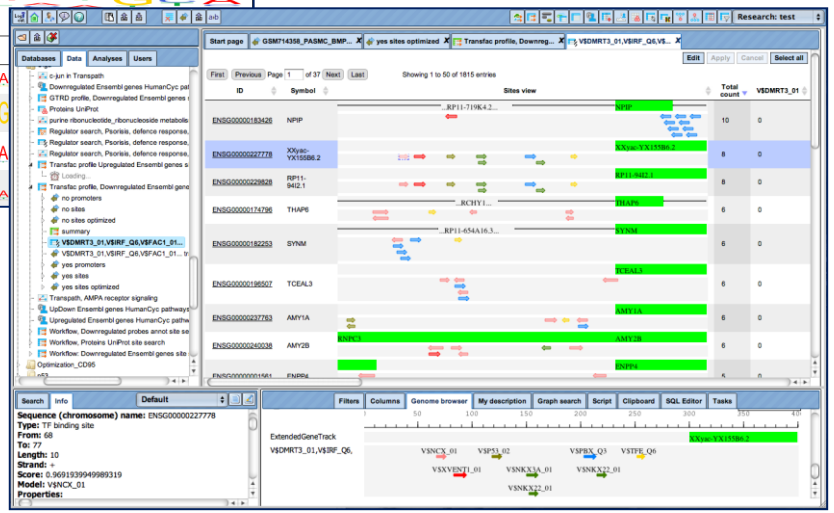
Binding sites referring to the same TF are merged into **positional weight matrices (PWM)**. A PWM reflects the frequency with which each nucleotide is found in each position of the known and aligned TFBSs and, thus, the base preference in each position.

Transcription factors are classified based on the general properties of their DNA-binding domains. The most up-to-date and comprehensive **TF classification** available has been included in the geneXplain platform.

Availablity

The most up-to-date version of TRANSFAC® can be obtained either

- for downloading as textual flat files, to have the full content locally at your disposal;
- for online access through the geneXplain platform, making full use of the rich functionality of this unique toolbox, or
- for online use of the familiar look-and-feel provided by a dedicated server, or
- any combination of these options.



Encyclopedic use

TRANSFAC® ist the most comprehensive encyclopedia about eukaryotic transcription factors. The structural and functional properties of each factor are documented by extensive manual annotation from the scientific literature by the BIOBASE team.

Individual TFBSs are documented including experimental details and a corresponding quality assessment.

Overview of high-throughput data

TRANSFAC® also documents HTP data on TF binding sites in eukaryotic genomes, usually from ChIP-chip or ChIP-seq experiments. These data are carefully selected and interpreted w.r.t. the binding regions and motifs found in the corresponding data sets.