

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® is the oldest and most comprehensive database on eukaryotic TFs. Since 1998, it has been taken over for further maintenance by BIOBASE GmbH, and was merged later on with the resources 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.

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.

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 to provide an integrated and comprehensive workflow management of a large number of "bricks", each providing a specific function in analyzing biological data
- In Silico Molecular Cloning (IMC) for handling large-scale genome data
- GenomeTraveler (GT) for handling next generation sequencing (NGS) data
- 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®

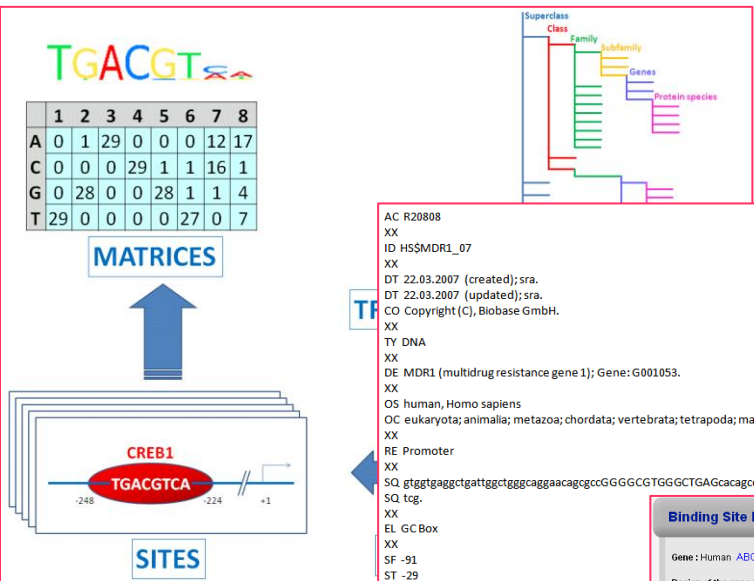
in the
geneXplain platform

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

BIOBASE
BIOLOGICAL DATABASES

geneXplain

TRANSFAC®: the database on eukaryotic transcription



Key features

- 34,000+ 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
- 18,000+ transcription factor reports (including miRNAs), a subset of which provide GO functional assignments, disease associations and expression pattern assignments
- 2,300,000+ ChIP fragment reports that include the best scoring site prediction for the respective factor as well as downloadable sequences and gene lists
- 277,000+ 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 1600+ 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.

Availability

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 use 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 the BIOBASE server, or any combination of these options.

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.

Encyclopedic use

TRANSFAC® is 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.

AC R20808
 XX
 ID H5\$MDR1_07
 XX
 DT 22.03.2007 (created); sra.
 DT 22.03.2007 (updated); sra.
 CO Copyright (C), Biobase GmbH.
 XX
 TY DNA
 XX
 DE MDR1 (multidrug resistance gene 1); Gene: G001053.
 XX
 OS human, Homo sapiens
 OC eukaryota; animalia; metazoa; chordata; vertebrata; tetrapoda; mammalia; eutheria; primates
 XX
 RE Promoter
 XX
 SQ gtggtaggctgattgctggcaggaacagcgcGGGGCGTGGGCTGAGcacagccgct
 SQ tgc.
 XX
 EL GCBox
 XX
 SF -91
 ST -29
 XX
 BF T00759; Sp1; Quality: 3; Species: human, Homo
 XX
 SO 0130; KB-3-1.
 SO 3658; KB-3-1+UV.
 XX
 MM direct gel shift
 MM functional analysis
 MM supershift (antibody binding)
 CC This site is required for UV irradiation-depende
 XX
 DR TRANSPRO:HSA_12513_1.
 XX
 RN [1]; R0049631.
 RX PUBMED: 10644769.
 RA Hu Z., Jin S., Scotto K. W.
 RT Transcriptional activation of the MDR1 gene by
 RL J. Biol. Chem. 275:2979-2985 (2000).
 XX
 //

Binding Site Information

Gene: Human ABCB1 (ATP-binding cassette, sub-family B (MDR/TAP), member 1)

Region of the gene: promoter

Sequence: gtggtaggctgattgctggcaggaacagcgcGGGGCGTGGGCTGAGcacagccgct FASTA

Sequence type: DNA

Position (relative to TSS unless stated): from -91 to -29

Description: GC Box

Promoters that the element is mapped to:
 ATP-binding cassette, sub-family B (MDR/TAP), member 1

Binding factors (with assigned measure):
 Sp1(1) Quality:3

Experimental source of the factors:
 KB-3-1; Human, epidermoid carcinoma cell line; KB-3-1+UV; Human, epidermoid carcinoma cell line

Method which measured binding: functional analysis

Identifiers

BIOBASE accession: R20808
 External accessions: EMBL/GenBank

Annotations

This site is required for UV irradiation-depende

References (1 entry)

[1] PMID 10644769 Hu Z., Jin S., Scotto K. W. J. Biol. Chem. 275 (4): 2979-85 (2000).

ENS0000022928

V5GR_Q6_01 8 GR

V5GSC_01 17 Gsc

V5GSH2_01 16 GSH2

V5GT2IRD1_01 9 GTF2IRD1-isoform2

V5GZF1_01 12

V5HAND1E47_01 16

V5H824_01 15

V5H89_01 16

V5HBP1_Q2 9

V5HDX_01 17 Hdx

V5HEB_Q6 6 HEB

V5HELOSA_01 11 Helos A

V5HELOSA_02 11 Helos A

Start page | GEM1438_PA3MC_BMP... | yes sites optimized | Transfac profile, Downreg... | V5DMRT3_01.V5RF_Q4.V5... |

Search: info (chromosome) Default

Sequence (chromosome) name: ENS0000022778

Type: TF binding site

From: 68

To: 77

Length: 10

Strand: +

Score: 0.969193994989319

Model: V5MCK_01

Properties:

ExtendedGeneTrack

V5DMRT3_01.V5RF_Q6

V5MCK_01 V5P51_Q2 V5P8X_Q3 V5P7E_Q6

V5XVNT1_01 V5MCK_01 V5MCK_01

V5MCK_01