

TRANSFAC® release 2019.2

The TRANSFAC® database on transcription factors, their genomic binding sites and DNA-binding motifs (PWMs), contains these new data features:

- **Matrices for methylated binding motifs**

1,785 matrices for human transcription factors from Methyl-HT-SELEX and HT-SELEX experiments were added. They were published in "Yin, Y. et al, Impact of cytosine methylation on DNA binding specificities of human transcription factors, Science 2017 356(6337):eaaj2239, Pubmed PMID: [28473536](#)".

Based on the selection of representative methylated motifs, a specific matrix profile has been added to be used with MATCH.

- **Drosophila melanogaster promoters added**

19,152 promoters for fruit fly (*Drosophila melanogaster*) have been included based on data from [Ensembl](#) version 95 and our [established](#) virtual transcription start site calculation. Experimentally verified transcription factor binding sites and 3.6 million SNPs from the dbSNP database have been mapped to the promoter sequences.

- **Annotation of transcription factor binding sites based on sequence conservation**

Known transcription factor binding sites located in human, mouse, rat, or pig genomes were extracted from TRANSFAC® and highly conserved sites were retained. Given high conservation as a prerequisite, binding sites were annotated for the three other species in respective genomic location if not more than one mismatch was observed in the sequence alignment with the primary species. This resulted in 1,007 new binding site entries.

- **Integration of new human and fruit fly ChIP-Seq experiments from ENCODE, modENCODE and modERN**

21 new human transcription factor binding site ChIP-Seq experiments released by the [ENCODE phase 3 and 4 project](#) between July 2018 and December 2018 have been integrated. The data sets comprise 424,961 fragments bound by 15 distinct transcription factors.

For *Drosophila melanogaster*, 470 transcription factor binding site ChIP-Seq experiments released by the [modENCODE](#) and [modERN](#) were added. The data sets include 2,050,835 fragments bound by 411 distinct transcription factors, all of which were not yet covered by TRANSFAC ChIP-Seq data.

For 183 of the sets, an existing positional weight matrix for the respective transcription factor was used together with the MATCH tool to predict altogether 953,185 best binding sites inside the fragments.

- **Ensembl version update**

Genomic information for genes, promoters, and ChIP fragments for the species human, mouse, rat, pig, macaque, *Drosophila*, and *Arabidopsis* is now based on Ensembl release 95.