

## TRANSFAC® release 2018.2

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

- **Performance assessment of TRANSFAC® PWMs and derived matrix recommendations**

Out of the huge collection of PWMs in the TRANSFAC database, a non-redundant library was compiled comprising the best-performing DNA-binding motifs of altogether 2799 transcription factors.

The user can now choose among four new PWM profiles consisting of recommended matrices for vertebrate, plant, fungal, and insect factors to be used with MATCH (to predict transcription factor binding sites, TFBSs, in DNA sequences) or FMATCH (to identify enriched TFBSs in a set of DNA sequences).

- **Integration of new human ChIP-Seq experiments from ENCODE**

164 new human transcription factor binding site ChIP-Seq experiments released by the [ENCODE phase 3 project](#) between October 2017 and January 2018 have been integrated.

The data sets comprise 2,570,897 fragments bound by 122 distinct transcription factors, of which 68 factors were not yet covered by ChIP-Seq data.

For 76 of the sets, an existing positional weight matrix for the respective transcription factor was used together with the MATCH tool to predict altogether 1,497,691 best binding sites inside the fragments.

Predicted best binding sites as well as complete fragments are available in FASTA and BED format via the ChIP Experiment Reports, as are lists of genes in a distance range to the fragments as specified by the user.

- **Addition of public human ChIP-Seq experiments from other sources**

1,757 human ChIP-Seq data sets published in [GEO](#) and [ArrayExpress](#) and re-analyzed by the [ReMap 2018](#) project have been incorporated. The experiments involve 48,509,720 fragments bound by 342 distinct transcription factors, including 190 without previous ChIP-Seq data set in the database. The peaks were taken from the “all peaks” catalog, allowing to preserve the cell specificity of the original experiments.

- **Ensembl version update**

Genomic information for genes, promoters, and ChIP fragments for the species human, mouse, rat, macaque, and Arabidopsis is now based on Ensembl release 91.