

TRANSFAC® release 2021.1

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

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

415 new human transcription factor binding site ChIP-Seq experiments released by the [ENCODE phase 4](#) project between June 2020 and September 2020 have been integrated. The data sets comprise 4,768,755 fragments bound by 391 distinct transcription factors, of which 214 factors were not yet covered by ChIP-Seq data in TRANSFAC.

For 336 of the sets, an existing positional weight matrix for the respective transcription factor was used together with the MATCH tool to predict altogether 3,652,932 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.

- **New matrices derived from ENCODE ChIP-Seq data**

128 new positional weight matrices have been generated from new ENCODE phase 4 ChIP-Seq data and integrated into the TRANSFAC matrix library.

- **JASPAR 2020 matrix library integration**

New position frequency matrices from the [JASPAR 2020](#) release either added as matrix entries (200 cases) or hyperlinked to existing counterparts in the TRANSFAC matrix library.

- **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 101.