TRANSFAC® release 2021.2

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

- **TRANSFAC 2.0 : Introducing the MATCH Suite**

  This software package has been optimized to identify transcription factors and their combinations that regulate the genes of your interest. MATCH Suite accepts human gene lists as input (either your own gene lists uploaded to the system or genes selected in the search results from the TRANSFAC®, TRANSPATH® or HumanPSD™ databases) and initializes a fully automatized sophisticated workflow, which produces the static analysis report and interactive tables of found factors and their binding sites. On-the-fly filtering by different criteria as well as the interactive genome browser visualization of sites identified in the promoters of the input genes is provided by the system in a user-friendly interface. MATCH Suite considers the biological context of your data, including functional categories or tissue specificities of your input genes, and provides you with the information on transcription factors and their combinatory modules that were identified to be regulating your input gene set.

- **New enhancer reports**

  199,183 human enhancer sequences have been imported from FANTOM5 and the HACER database and lifted over to the GRCh38 genome assembly. The enhancer reports display genes with which promoters the enhancer interacts, tissues and cell types/lines the enhancer is active in, and genomic regions such as histone modification sequences, DNase I hypersensitivity sites, and transcription factor binding sites that overlap with the enhancer.

- **Additional transcription factor interactions**

  3,888 new interactions between human transcription factor proteins from, among others, the recently published BIOPLEX 3 data set and the human reference interactome (HuRI) have been included.

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

  15 new human transcription factor binding site ChIP-Seq experiments released by the ENCODE phase 4 project have been integrated. The data sets comprise 197,893 fragments bound by 13 distinct transcription factors, of which 5 factors were not yet covered by ChIP-Seq data in TRANSFAC.
  For 13 of the sets, an existing positional weight matrix for the respective transcription factor was used together with the MATCH tool to predict altogether 174,539 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.

- **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 102.