TRANSFAC® release 2017.2

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

- 528 new matrices (DNA-binding motifs) for Arabidopsis thaliana transcription factors based on DAP-Seq (DNA affinity purification sequencing) (O'Malley et al., 2016 Cell 165: 1280-1292 https://www.ncbi.nlm.nih.gov/pubmed/?term=27203113) and 46 matrices for Arabidopsis thaliana transcription factors based on PBM (protein binding microarrays) (Sullivan et al., 2014 Cell Rep. 8:2015-2030 https://www.ncbi.nlm.nih.gov/pubmed/?term=25220462).
- 521 DAP-Seq data sets from the Arabidopsis cistrome database comprising 2,826,587 fragments (O'Malley et al., 2016 Cell 165: 1280-1292 https://www.ncbi.nlm.nih.gov/pubmed/?term=27203113) analyzed with the MATCH tool for 2,682,764 best binding sites inside the fragments.
- 253 new transcription factor binding site ChIP-Seq experiments released by the ENCODE phase 3 project between October 2016 and January 2017 https://www.encodeproject.org/matrix/?type=Experiment&status=released. The data sets comprise 4,622,342 fragments bound by 213 distinct transcription factors, of which 157 factors were not yet covered by ChIP-Seq data. For 142 of the sets, an existing positional weight matrix for the respective transcription factor was used together with the MATCH tool to predict altogether 2,841,686 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.
- Genomic information for genes, promoters, and ChIP fragments for the species human, mouse, rat, macaque, and Arabidopsis is now based on Ensembl release 87.
- New November 2016 dbSNP data releases for human and rat have been integrated.

PROTEOME™ (HumanPSD™+TRANSPATH®) release 2017.2

The Human Proteome Survey Database (HumanPSD™) with focus on human proteins as disease biomarkers and drug targets contains these new data features:

- Integration of new data on clinical trials (https://clinicaltrials.gov/) (Increase of CT-Disease-Drug assignments from 146,605 to 227,170 due to inclusion of new data as well as improved mapping)

The TRANSPATH® database on mammalian signal transduction and metabolic pathways contains these new data features:

- 4,589 additional miRNA family entries according to TargetScanHuman release 7.1 (http://www.targetscan.org/vert_71/), allowing the transformation of 24,488 miRNA–target gene reactions from experimental evidence to semantic level