

## geneXplain® platform 6.3 release

### Database updates:

- ✕ HumanPSD™ is updated to version 2021.1 (January 2021).
- ✕ TRANSFAC® is updated to version 2021.1 (January 2021).
- ✕ TRANSPATH® is updated to version 2021.1 (January 2021).

### New workflow:

#### ✕ **Identify enriched motifs in cell specific promoters (TRANSFAC(R))**

This workflow searches for enriched transcription factor binding sites (TFBSs) in a set of gene promoters versus a random promoter set. The input gene set is used to extract promoter regions by mapping it against the TSS locations defined in the Fantom5 (Nature 507:462–470) database for one selected cell-type among 172 available cell-types. The over-represented sites identified with the MEALR method are converted into a profile, which is used for a second round of site search analysis and ends up with the identification of potential transcription factors.

### New features:

#### ✕ **MEALR classifier (tracks)**

MEALR searches for a combination of transcription factor binding motifs that discriminate between a positive (Yes) and a negative (No) sequence set. This tool takes a sparse logistic regression model derived with MEALR and applies it to new sequences to predict whether they can be bound by TF complexes or contribute to gene regulation in the same way as the Yes sequences used to train the MEALR model.

#### ✕ **Random forest prediction**

Random forests are a combination of tree predictors and that each tree depends on the values of a random vector sampled independently and with the same distribution for all trees in the forest. This statistical method performs a classification or

regression with a random forest model, based on Breiman (published in [Machine Learning](#)). Please refer to documentation of the R [randomForest](#) package for computational details. Random forests can be trained using the Train random forest tool (see below).

#### ✕ **Train random forest**

This method can be used to train a random forest model for classification, regression, or clustering. Please refer to documentation of the R [randomForest](#) package for computational details. Except for unsupervised models, the trained random forests can be used for further classification or regression analysis with the Random forest prediction tool.

#### ✕ **t-SNE**

The new tool contains a R wrapper around the fast T-distributed Stochastic Neighbor Embedding implementation by Van der Maaten (more information on the original implementation is [here](#)). The tool can be used for data visualization using the t-SNE algorithm.

### Enhancements:

- ✕ **Fantom5 workflows now available for both, hg19 and hg38 genome versions.**
- ✕ **The workflow for analyzing a SNP list with TRANSFAC database is now available for the hg38 genome version.**
- ✕ **New example with Gene Expression Omnibus data (GSE156063): Upper airway gene expression differentiates COVID-19 from other acute respiratory illnesses and reveals suppression of innate immune responses by SARS-CoV-2, Expression profiling by high throughput sequencing, Illumina NovaSeq 6000 Homo sapiens.**