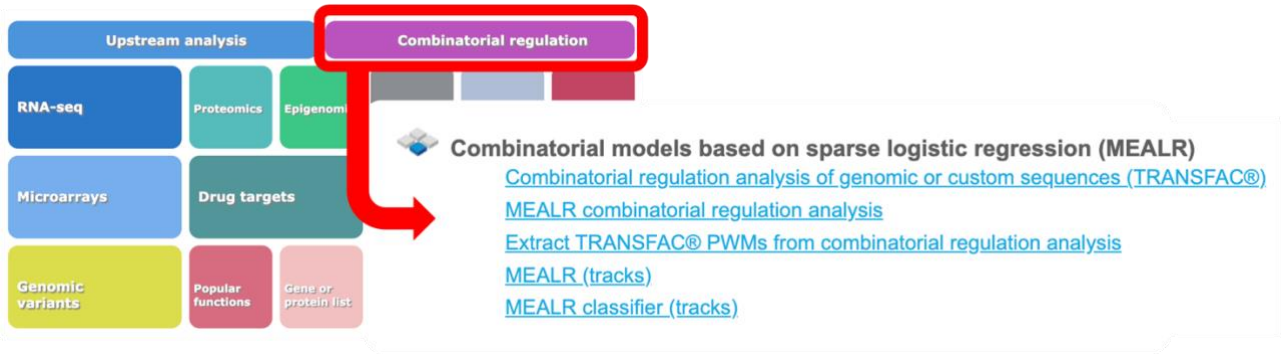
July 2023

geneXplain® platform 7.2 release

New methods and workflows:

This release of the geneXplain platform introduces a new section of the start page called "Combinatorial regulation", which provides tools for analysis of combinations of TF binding sites in promoters, enhancers and silencers, and other regulatory genomic regions. In this section, in addition to our traditional Composite Module Analyst, we introduce a completely new branch of site search combinatorial analysis methods based on the sparse logistic regression (MEALR):



The MEALR subsection includes the following new workflow:

Combinatorial regulation analysis of genomic or custom sequences

This workflow scans input sequences for tissue and cell type- specific transcription factor binding regions using the newly created library of MEALR models delivered with the TRANSFAC® database. (The library includes MEALR models for more than 1000 transcription factors trained on ChIP-seq data on TF binding in specific cell type and cell type condition).

Track of genomic intervals or FASTA file with a custom sequence can be used as input. Cell and/or tissue source can be specified for customization of the search results.

Prediction of functional transcription factor (TF) binding sites is a difficult task, because recognized DNA elements are rather short and often do not follow simple rules with regard to sequence specificity. Formation of TF-DNA complexes depends on a context determined by intertwining conditions like cellular differentiation, chromatin state, or expression and activity of cooperating TFs. To overcome these difficulties and produce reliable predictions of transcription factor binding sites in the scanned sequences, we have constructed a TRANSFAC® library of MEALR models providing the first comprehensive collection of tissue and cell type-specific TF binding models that account for combinatorial TF-DNA complexes comprising multiple DNA-binding specificities as well as cellular and tissue-related contexts. The library includes models of binding of almost 1000 human TFs in over 400 cell types and more than 50 tissue types. These models deliver predictions with high accuracy. The internal structure of the TF binding region is analyzed in a subsequent search for TF binding sites using relevant positional weight matrices (TRANSFAC® PWMs) that mark the location of binding of the TF in focus and locations of binding of the specific combination of co-factors that accompany the TF binding in particular tissue and cellular conditions. As a result, within the predicted long TF binding regions, MEALR model furthermore suggest several binding sites for TFs that in their specific combination accompany the TF of interest and provide the tissue and cell type specificity of TF binding.

You will find a more detailed description of this new workflow in the geneXplain platform user guide.

As well as the following new geneXplain platform methods:

MEALR combinatorial regulation analysis

This method applies combinatorial regulatory models (CRMs) based on the MEALR affinity score *[Lloyd, Katie, et al. "Using systems medicine to identify a therapeutic agent with potential for repurposing in inflammatory bowel disease." Disease models & mechanisms 13.11 (2020): dmm044040]* to classify or scan sequences for occurrences of combinations of transcription factor binding sites represented by TRANSFAC® PWMs. The models are taken from the MEALR library whose training data originate from the TRANSFAC® collection of high-throughput ChIP-seq sequencing experiments.

This method can be launched in a classification or in a scan mode. *Classification mode* evaluates input sequences as a whole, whereas the *scan mode* analyzes sequence windows separated by the given *step size* (sliding window). Cell and tissue sources can be selected to focus on a subset of CRMs which have been trained with data from respective sources.

You will find a more detailed description of this new method in the geneXplain platform user guide.

Extract TRANSFAC(R) PWMs from combinatorial regulation analysis

This tool extracts TRANSFAC® PWMs from a result table generated by the *MEALR combinatorial regulation analysis*. The PWMs represent transcription factor binding motifs for the TF in focus and its co-factors that constitute the combinatorial module predicted by the MEALR model.
You will find a more detailed description of this new method in the geneXplain platform user guide.

New import possibility:

The files from the Sequence Read Archive (SRA) can be now uploaded to the geneXplain platform directly by the SRR ID. You will find a detailed description of the new import possibility in the geneXplain platform user guide.

Database updates:

- HumanPSD™ is updated to version 2023.1 (July 2023).
- TRANSFAC® is updated to version 2023.1 (July 2023).
- TRANSPATH® is updated to version 2023.1 (July 2023).