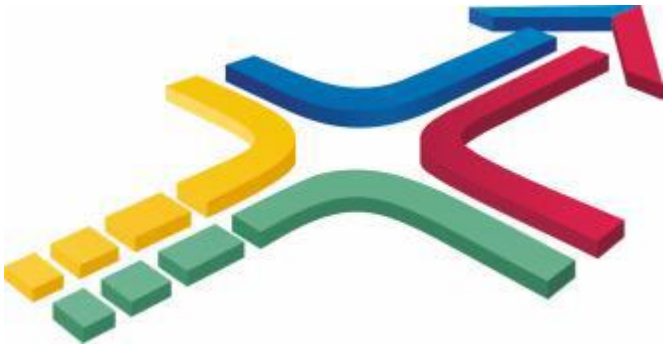
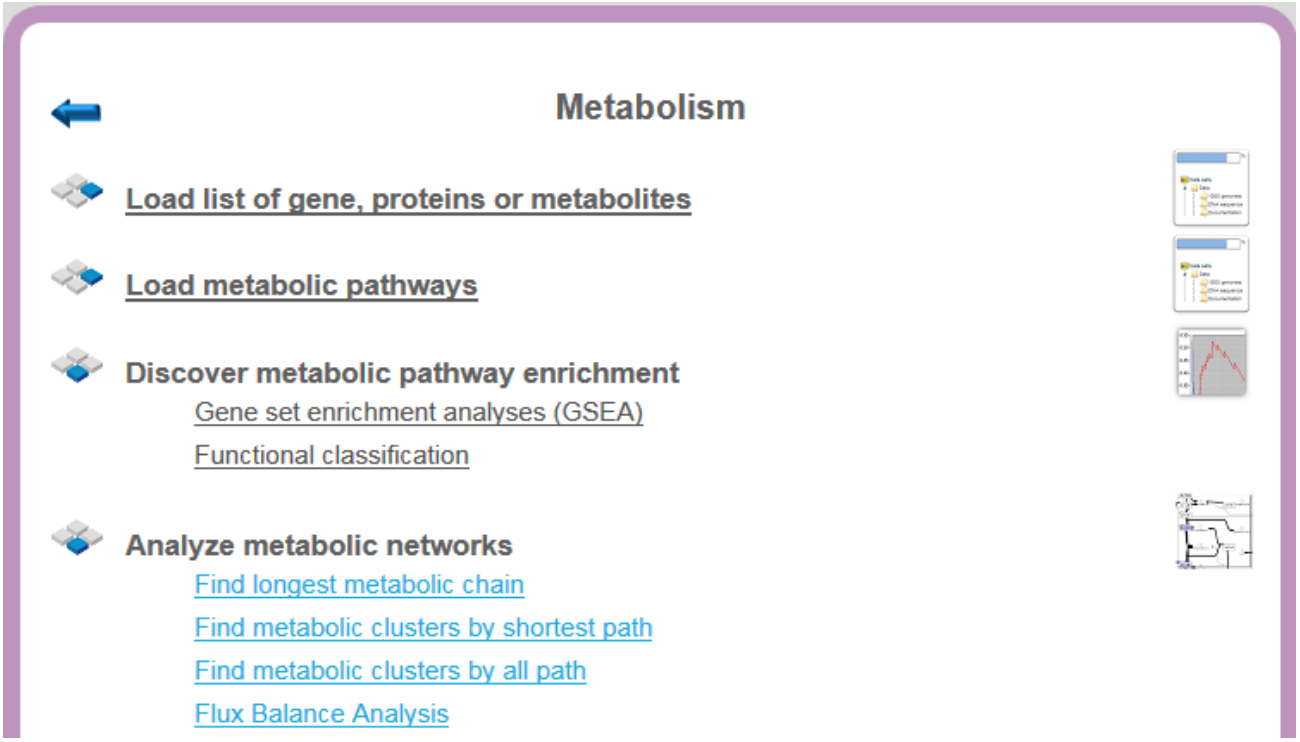
New features of the geneXplain platform release 4.4, January 2017



1. New area on start page: Metabolism

Load a list of genes, proteins, metabolites or metabolic pathways and discover metabolic pathway enrichment or analyze metabolic networks.



2. New integrated and updated databases

Human Metabolic Reaction (HMR)

The Human Metabolic Reaction (HMR) is a genome-scale metabolic model for the generic human cell. In its latest version (HMR2), it contains 3.765 genes associated with over 8.000 reactions and over 3.000 unique metabolites.

http://www.metabolicatlas.org/downloads/hmr

Genome-scale metabolic modelling of hepatocytes reveals serine deficiency in patients with non-alcoholic fatty liver disease. Nat Commun. 5: 3083 (2016).

http://www.nature.com/ncomms/2014/140114/ncomms4083/full/ncomms4083.html

Human metabolism global reconstruction (Recon 2)

Recon 2 is a community-driven consensus 'metabolic reconstruction', which is the most comprehensive representation of human metabolism that is applicable to computational modelling.

For the 'reconstruction' of Recon 2, metabolic information present in four different resources (EHMN17, HepatoNet1, Ac-FAO module and the human small intestinal enterocyte reconstruction) was added to the content of Recon 1 database.

7440 reactions and 2626 different metabolites are available in the Recon 2 database and has been described in:

Thiele I, et al.: A community-driven global reconstruction of human metabolism. Nature Biotechnology 31(5): 419-425 (2013).

http://www.nature.com/nbt/journal/v31/n5/full/nbt.2488.html

Recon 2 is freely available at http://humanmetabolism.org/.

HOmo sapiens COmprehensive MOdel COllection (HOCOMOCO)

Partly based on TRANSFAC®, HOCOMOCO provides transcription factor (TF) binding models for 641 human and 427 mouse TFs.

http://hocomoco.autosome.ru/

HOCOMOCO provides position weight matrices (PWMs), based on ChIP-Seq data. All TF binding models were produced by the ChIPMunk motif discovery tool.

Kulakovskiy et al.: HOCOMOCO: a comprehensive collection of human transcription factor binding sites models. Nucleic Acids Res. 41: D195-202(2013).

http://nar.oxfordjournals.org/content/44/D1/D116

TRANSFAC® and **TRANSPATH®** updated to releases 2017.1

Reactome

The database provides information on biological objects such as proteins, protein complexes, reactions etc. and is updated to version 57.

Reactome57 contains 4388 proteins, 2197 protein complexes and 1810 reactions.

Newly added **Ensembl** genome builds:

Ensembl Homo sapiens (hg38)

Ensembl Mus musculus (mm10)

Ensembl Rattus norvegicus (rn6)

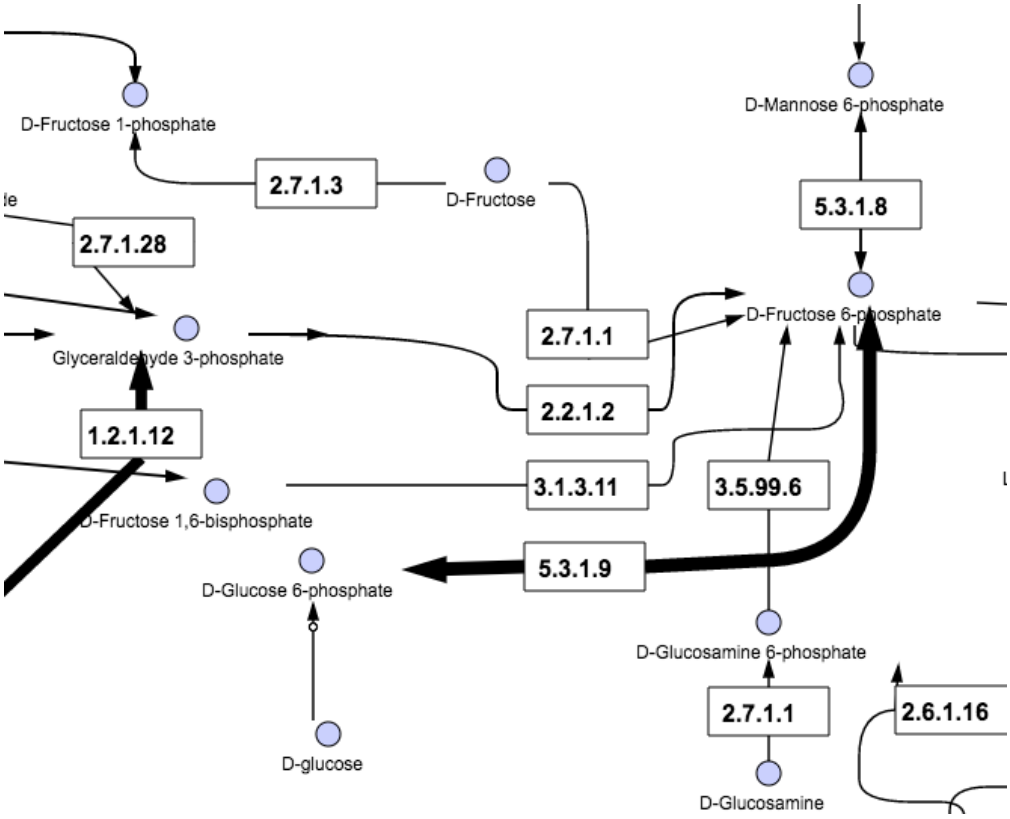
3. New workflows

Flux Balance Analysis

The workflow supports the identification of enzymes, reactions and metabolites.

Any input set of genes or proteins will be converted into a list of enzymes and matched with the content of the metabolic databases HMR or Recon2 to get a table with flux data. In parallel, the enzyme list is converted into an enzyme reaction table.

Finally all identified enzyme reactions are visualized with the corresponding EC numbers and named metabolites, a weighting of the reactions and the existence of metabolites is shown via the thickness of the reaction arrows (see example below).



Upstream analysis with feedback loop (TRANSFAC® and TRANSPATH®)

This workflow is the classical upstream analysis workflow with additional selection of the master regulators, which are differentially expressed. The output is the list of master regulators and the three best networks.

Search for self-regulating transcription factors (TRANSFAC® and TRANSPATH®)

This workflow comprises an upstream analysis with a focus on transcription factors (TFs). Input is a list of differentially expressed genes (DEGs). Output will be a list of TFs that are differentially expressed, **and** whose sites are overrepresented in promoters of the differentially expressed genes. The output is often a short list of transcription factors, found by the criteria to be involved the self-regulating loops.

Identify enriched motifs in tissue specific tracks (TRANSFAC(R))

This workflow is designed to search for overrepresented transcription factor binding sites (TFBS) in a sequence collection (track) with cell-specific or tissue-specific transcription start sites (Fantom5 data).

Cross-species identification of enriched motifs in promoters

The first step in this workflow is a conversion from mouse or rat genes into human orthologue genes, followed by a search for putative transcription factor binding sites (TFBS) in the promoters of these human genes.

Cross-species mapping to ontologies

The first step here is a conversion from mouse or rat genes into human orthologue genes. After that the human genes are mapped to ontology data.

4. New methods

Remove overlapping sites

This method removes overlapping sites from any track and constructs a subset of the input track with no sites overlap in the output track. Different modes construct different sets of non-overlapping sites (see example below):

One longest

One shortest

One with best value

One random

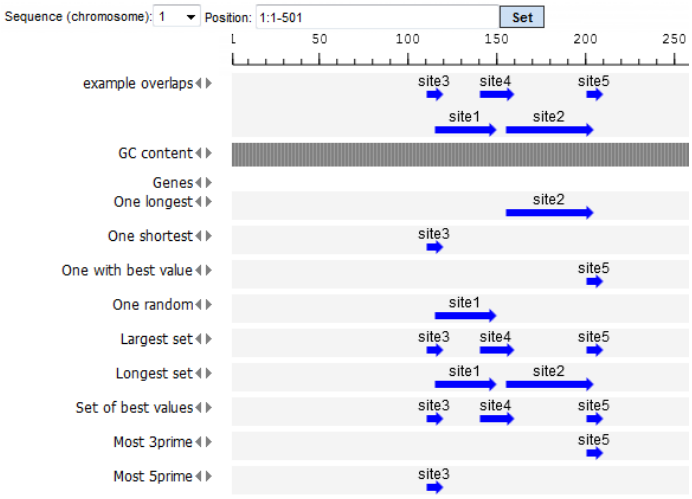
Largest set

Longest set

Set of best sites

Most 3prime

Most 5prime

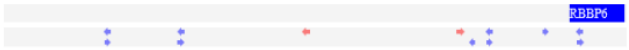


Preprocess raw reads

This new method can be used for a preprocessing step of single-end or pair-end fastq or solid csfastq files to remove reads with low quality or adapters as well to trim low quality bases from read ends.

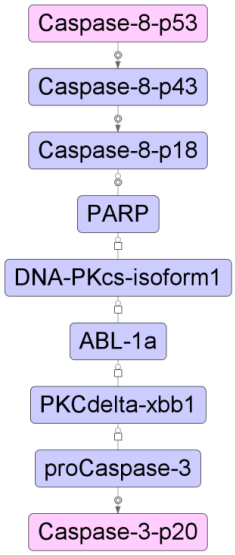
Site search report

This method generates a visualization of selected models with exact site positions and number of sites within the promoters from a site search summary. In the example below on the left side two gene promoters are shown with sites (= arrows in the column **Sites view**) from three selected site models (red, blue and turquoise). The **Total count** of sites for the first gene promoter is 44. The remaining columns of the right side give the individual number of each site and the position(s) in the gene promoter.

Sites view	Total count	V$HBP1_03	V$CPHX_01	V$MAZR_01	V$HBP1_03 positions	V$CPHX_01 positions	V$MAZR_01 positions
	44	17	27	0	-873, -691, -682, -678, -670, -634, -577, -536, -524, -520, -480, -438, -421, -353, -302, -298, -294	-906, -905, -866, -865, -830, -829, -821, -778, -777, -703, -692, -691, -678, -675, -674, -662, -... (more)	
	12	2	10	0	-473, -200	-823, -822, -693, -692, -178, -148, -147, -48, 12, 13	

Find longest connected chains

The goal of this analysis is to find the longest chains which contains as many elements from an input collection as possible. Here chain means a path which starts and ends with the elements from the input collection. The path length between two elements from the input collection is limited by the maximal search radius.



Convert site search summary

Input is a table from a transcription factor binding site search and at least one column with numerical values (e.g. p_value). The user selects one of the rules to treat values in the numerical columns of the table when several rows are merged into a single one.

The possible rules are:

average, average w/o 20% outliers, sum, minimum, maximum or extreme

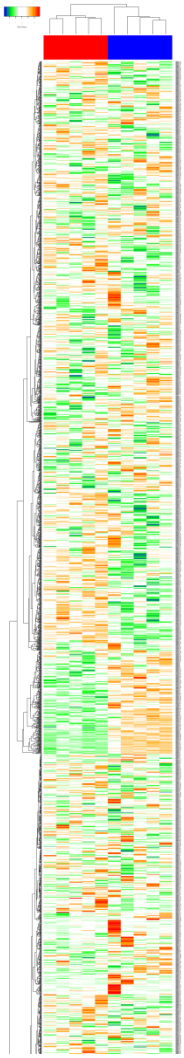
The chosen rule is applied to all numerical columns of the table. Output is the converted table with the rule applied.

Heatmap

This tool creates a heatmap for the numerical data matrix provided with the input data table.

The output folder contains a TIFF image of the heatmap as well as the ordered lists of row ids (e.g. RNA or gene ids) and column ids.

The output tables can be used to extract subsets of correlated rows or columns revealed by the hierarchical clustering and/or the heatmap presentation.



Variance filter This tool creates two tables with the specified number of rows with highest and lowest variance, respectively.