geneXplain® platform 6.2 release

Database updates:

- Ensembl is updated to release 100 (APRIL 2020).
- Reactome database is updated to version 74 (September 2020).
- HumanPSD™ is updated to version 2020.3 (October 2020).
- TRANSFAC® is updated to version 2020.3 (October 2020).
- TRANSPATH® is updated to version 2020.3 (October 2020).

New workflow:

- **Compute differentially expressed genes using Limma and Metadata**

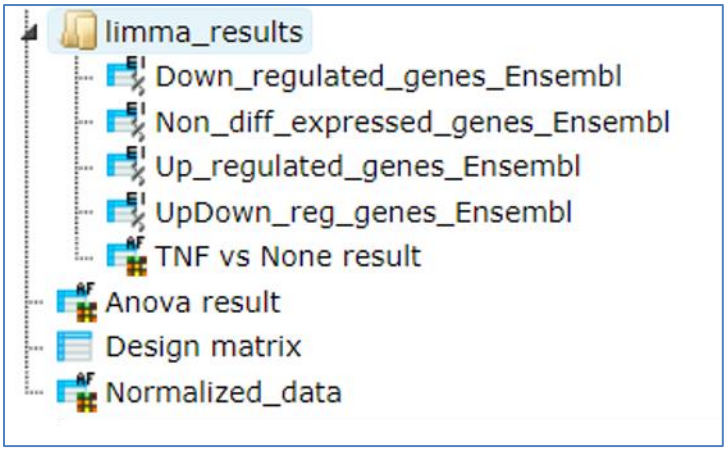
This workflow performs a linear model analysis to identify differentially expressed genes from multiple samples using Limma statistics and a metadata table for the samples. The given input table contains expression values from several samples and a corresponding sample table (metadata) for guiding the limma analysis by selected experimental factors. The workflow aims at finding significant differences between pairs of levels of a main factor (**Treatment**). Furthermore, an ANOVA is carried out for all contrasts together. The assignment of main factor levels to columns of the input table is specified in a column of a sample table.

Sample table may look like this:

Sample	Treatment	Cell	Batch	TreatCell
GSM50771.CEL	None	HMEC	A	NoneHMEC
GSM50772.CEL	None	HMEC	A	NoneHMEC
GSM50773.CEL	None	HMEC	B	NoneHMEC
GSM50774.CEL	TNF	HMEC	B	TNFHMEC
GSM50775.CEL	TNF	HMEC	C	TNFHMEC
GSM50776.CEL	TNF	HMEC	D	TNFHMEC
GSM50777.CEL	None	HUVEC	D	NoneHUVEC
GSM50778.CEL	None	HUVEC	C	NoneHUVEC
GSM50779.CEL	None	HUVEC	E	NoneHUVEC
GSM50780.CEL	None	HUVEC	F	NoneHUVEC
GSM50781.CEL	TNF	HUVEC	E	TNFHUVEC
GSM50782.CEL	TNF	HUVEC	F	TNFHUVEC
GSM50783.CEL	TNF	HUVEC	G	TNFHUVEC
GSM50784.CEL	TNF	HUVEC	G	TNFHUVEC

Raw counts will be processed using Limma's voom method, optionally including the specified normalization method, whereas Normalized expression values are used as is, and for Transformed counts an intensity-based trend is included during Limma analysis (eBayes parameter trend=TRUE).

The primary result of the linear model analysis is further filtered to identify significant up- and down-regulated genes for each sample comparison.



New features:

Join full tables

Joining two tables into a new one with containing the selected columns. Different joining types can be processed according to the ID matching from both input tables.

Columns with the same names can be merged if the respective option is checked. Otherwise column name will be extended by table name for both input tables and two distinct columns will be taken to the resulting table.

Following joining types are available:

- **Inner** - inner join in SQL sense (result will contain IDs presented in both tables).
- **Outer** - full outer join in SQL sense (result will contain IDs presented at least in one table).
- **Left** - left outer join in SQL sense (result will contain IDs presented in left table).
- **Right** - right outer join in SQL sense (result will contain IDs presented in right table).
- **Left subtraction** - result will contain IDs presented ONLY in left table.
- **Right subtraction** - result will contain IDs presented ONLY in right table.
- **Symmetric difference** - result will contain IDs presented only in one (any) of two tables.

- **Calculate keynodes ranks**

This method takes from the result of 'Regulator search' analysis the column of gene identifiers. It adds optional score columns from three other tables called *Transcriptome*, *Proteome* and *CMA score table*. Any table of gene identifiers with numerical score column can be set as any of these three tables.

The result is ranked via the Master Regulator Score (Rank sum) column and via the newly added scores. The ranking procedure used is the following: molecules are sorted in descending order of scores and then integer numbers starting from 1 are assigned to each molecule. Molecules with the same score obtain the same rank number. The Total Rank is the sum of score-specific ranks for each molecule to identify the best corresponding master regulator.

- **Select keynodes with top targets**

This method selects the top master regulators (keynodes) by score and linkage to top targets. As input the method uses the result of 'Regulator search' analysis and the target table of 'PSD pharmaceutical compounds analysis'. The method sorts the input table by Rank sum column and takes the number of top elements molecules from the sorted table. It adds more molecules from the top so that the union of column with gene IDs values will cover at least the number of top targets from the target table.

The maximum number of molecules that can be obtained is:

Number of top elements + Number of top targets * number_of_target_table

Enhancements:

- **Update MTB report to version 2.0.0**
- **Removal of redundant Ensembl versions for the same build**
- **Add Ensembl annotation source to all workflows**
- **Bug fixing of Affymetrix miRNA chips normalization**
- **Easy selection of current data project**