MetaboAnalyst 5.0

A Web-based Tool for streamlined Metabolomics Data Analysis

2022.07.12

7. Statistical Analysis with metadata

Statistical analyses [Metadata table] that take these covariates into account can draw more robust conclusions about the relationships between the primary variable and the omics data. Statistical Analysis with metadata become available now.

Highlights:

- This module is designed to perform covariate analysis, including covariate adjustment and identifying major patterns with regard to the two given factors and their interactions.

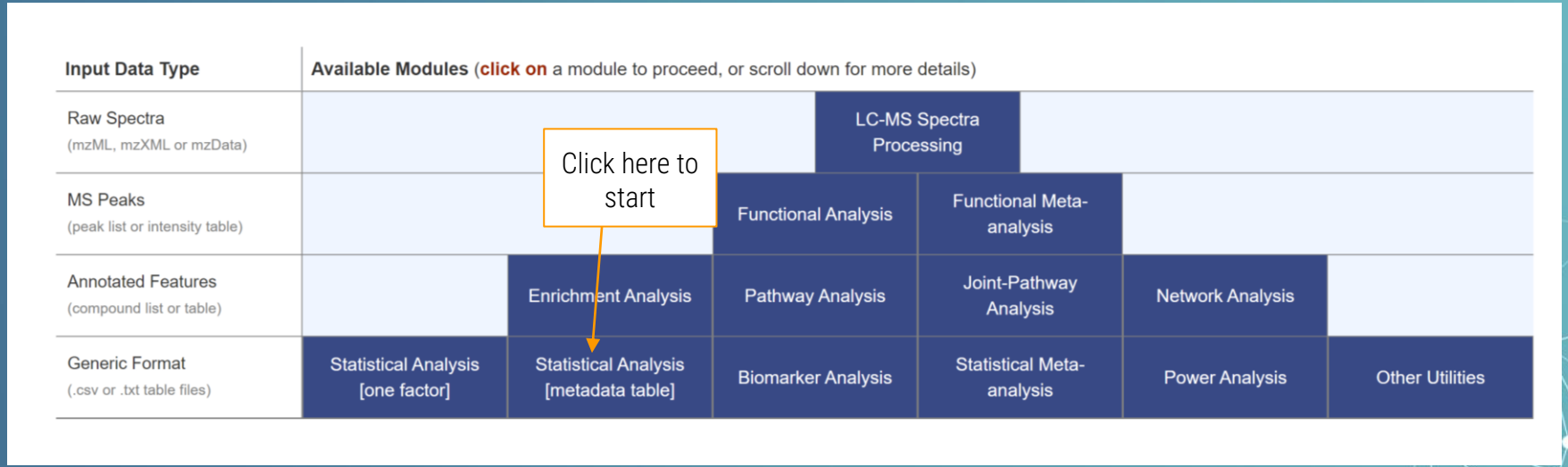
7.0 Knowledge & Background

Covariate Analysis

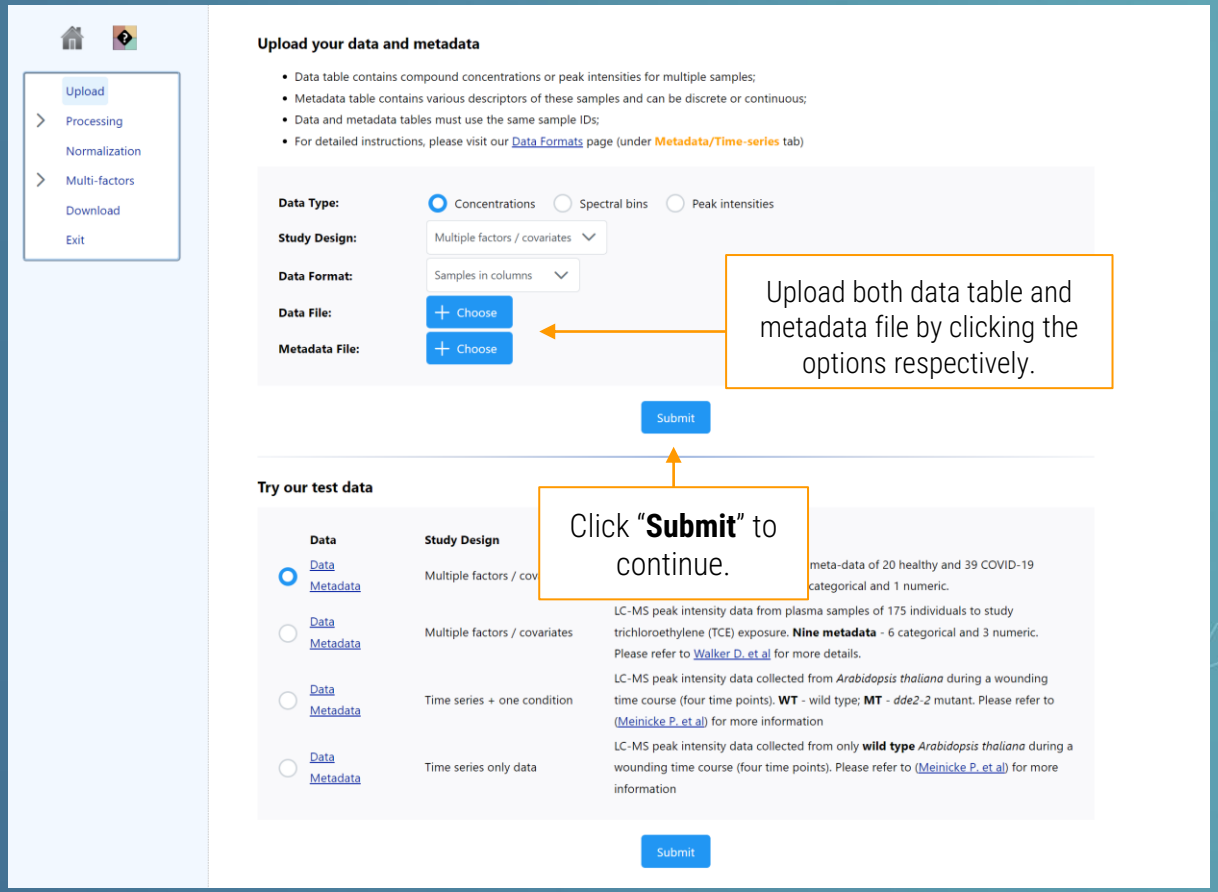
- Metadata describes the data, and typically contains details on the experimental condition (i.e., treatments), sample source (i.e., species, tissue), and sample collection (i.e., location, time). Such metadata are critical for data interpretation, allowing researchers to analyze the data with respect to their biological and environmental context, and for data re-use, allowing other researchers to search for, and meaningfully compare and potentially integrate, results from across diverse studies.

- Details on the environmental context and sample source are becoming increasingly important as observational studies that collect omics data from human populations or animals outside of laboratory settings are becoming more common43. In these cases, there is typically a primary variable of interest, such as presence/absence of a certain disease or exposure to a specific chemical, as well as other variables such as age, sex, or other potential factors that co-vary with the primary metadata

7.1 Start Statistical Analysis [Metadata Table]

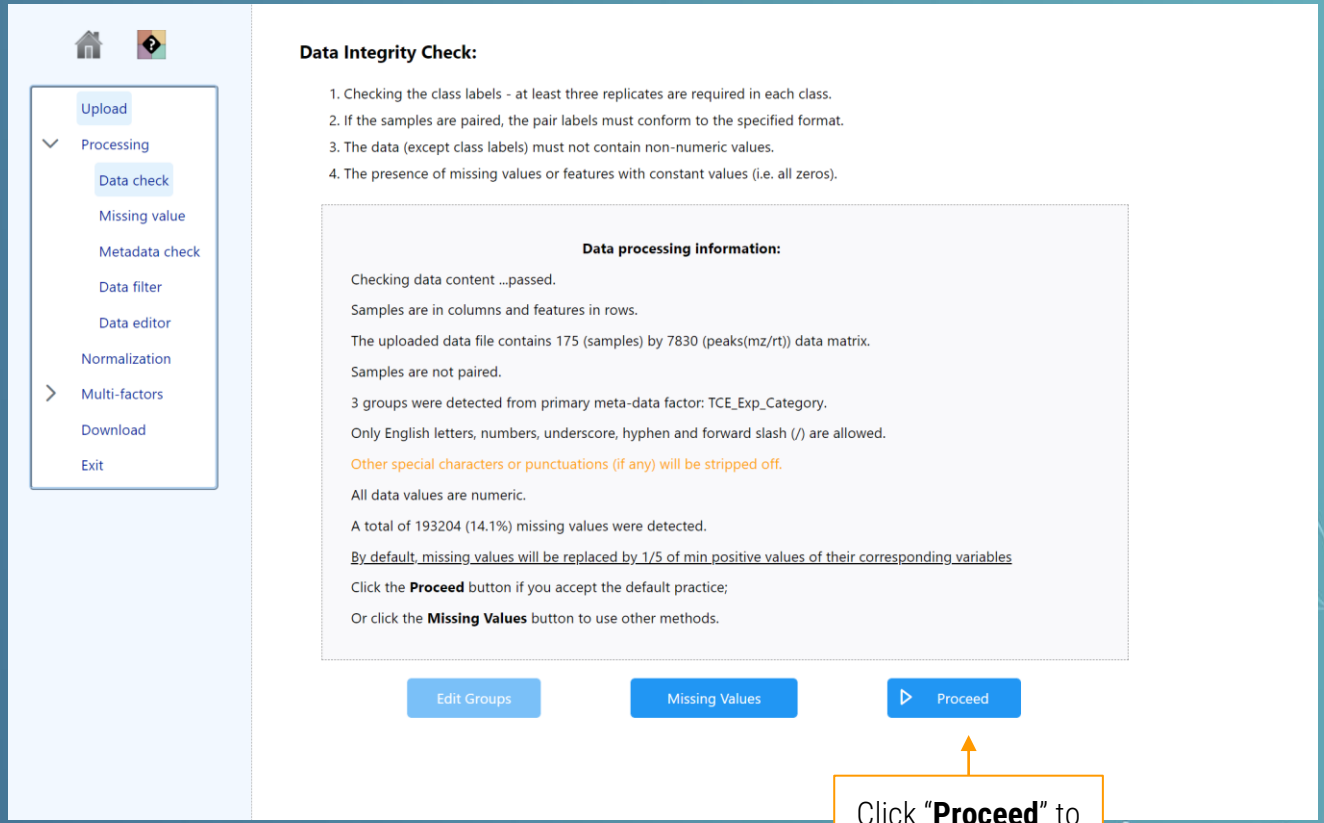


7.2 Data Upload Page

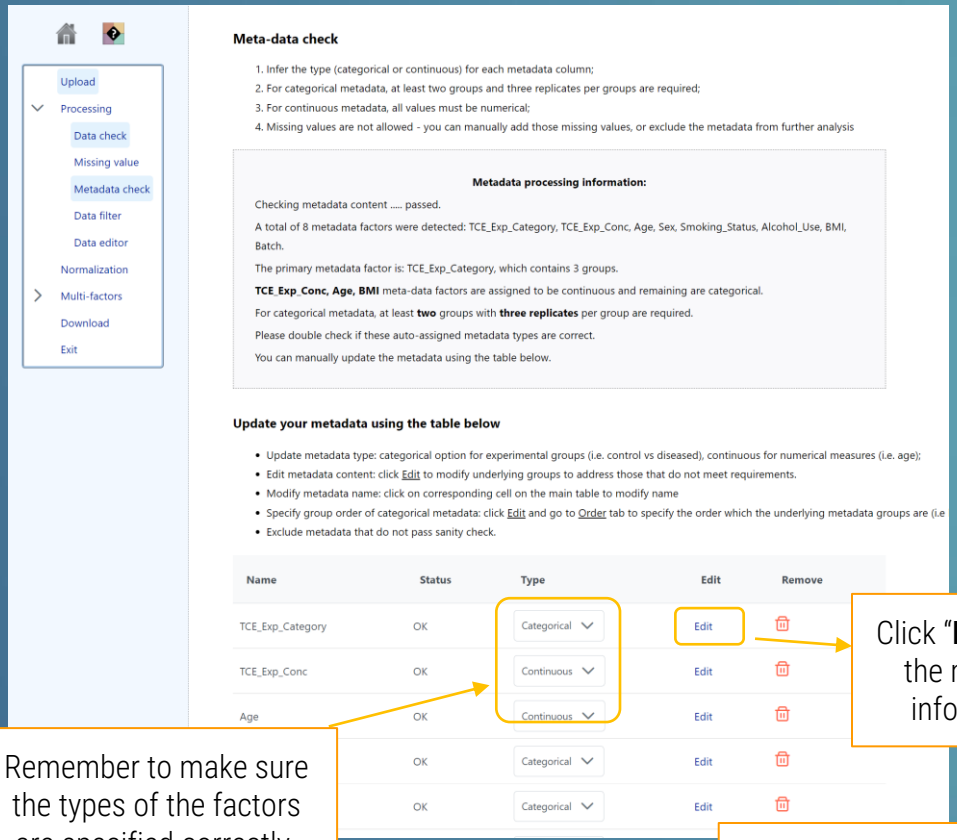


TIP: A separate metadata file is required for Statistical Analysis [metadata table], in addition to a data table containing feature abundance information. The first column of the metadata must be the same sample names used by the data file. The second column contains the primary condition of interest. Other metadata (such as sex, age, or batch) can be included in the remaining columns. Both continuous and discreate variables are acceptable. MetaboAnalyst cannot deal with missing values (NA or empty) in the metadata. Users will be asked to manually fix these values if detected.

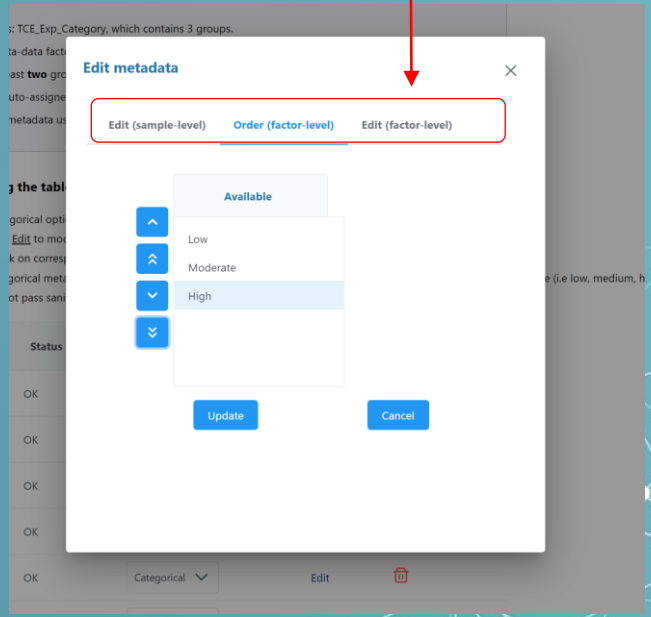
7.3 Data integrity check



7.4 MetaData integrity check



You can change the levels' order or name of the metadata.

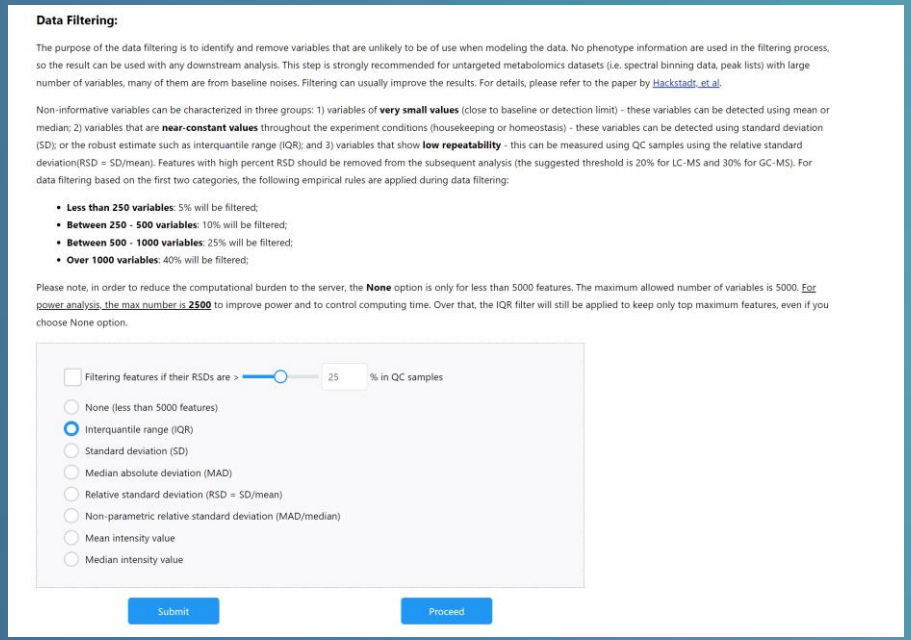


Click "**Edit**" to edit the metadata information.

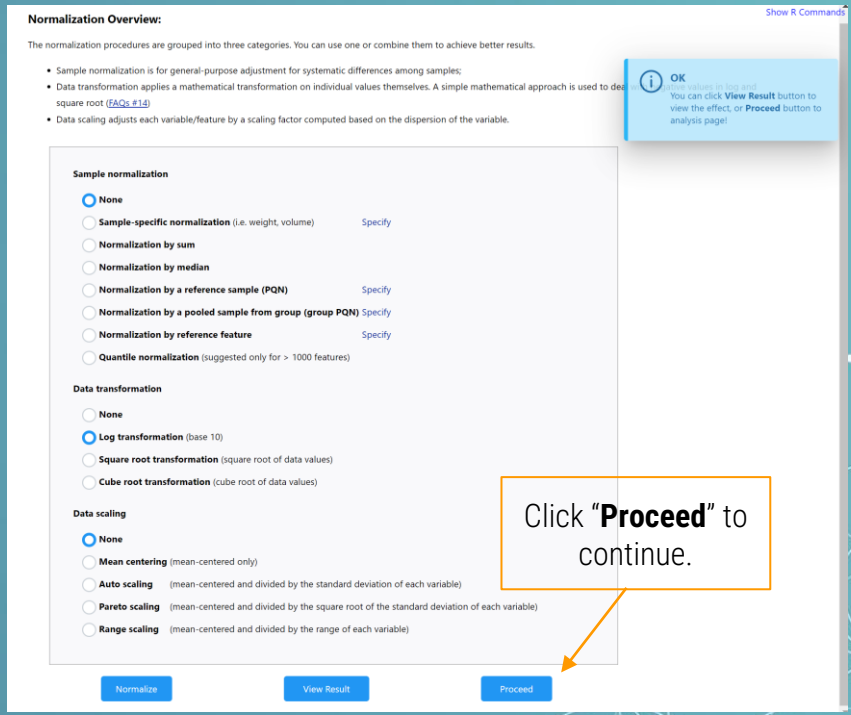
Remember to make sure the types of the factors are specified correctly.

Click "**Proceed**" button from the bottom to continue.

7.5 MetaData integrity check

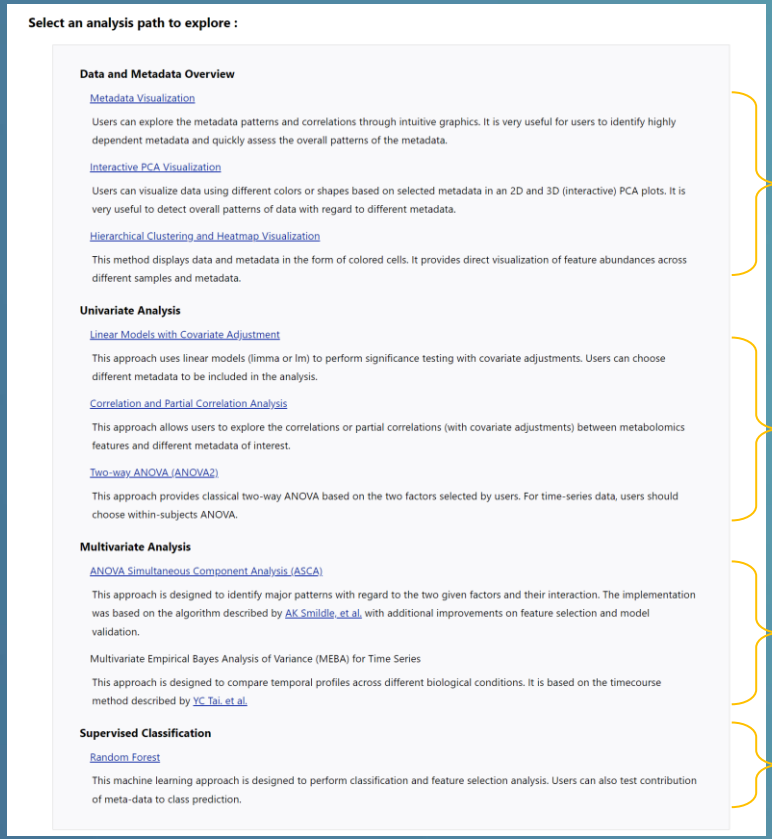


Click "**Proceed**" to continue.



Click "**Proceed**" to continue.

7.6 Selection of Analyzing Items



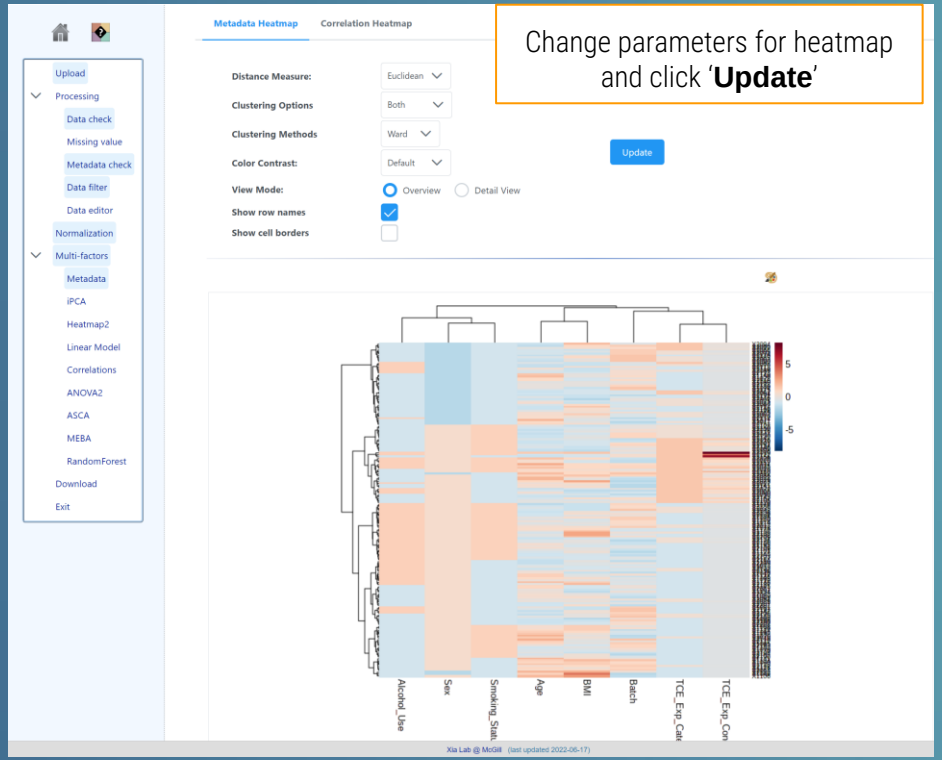
Data and Metadata Overview

Univariate Analysis

Multivariate Analysis

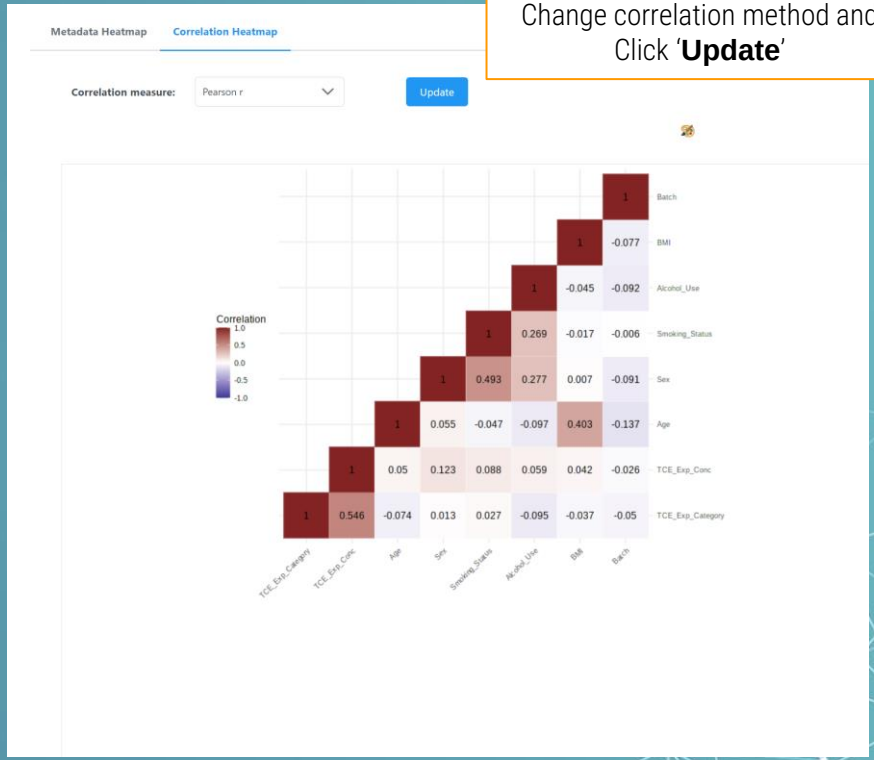
Supervised Classification

7.6 Metadata Visualization



Change parameters for heatmap and click '**Update**'

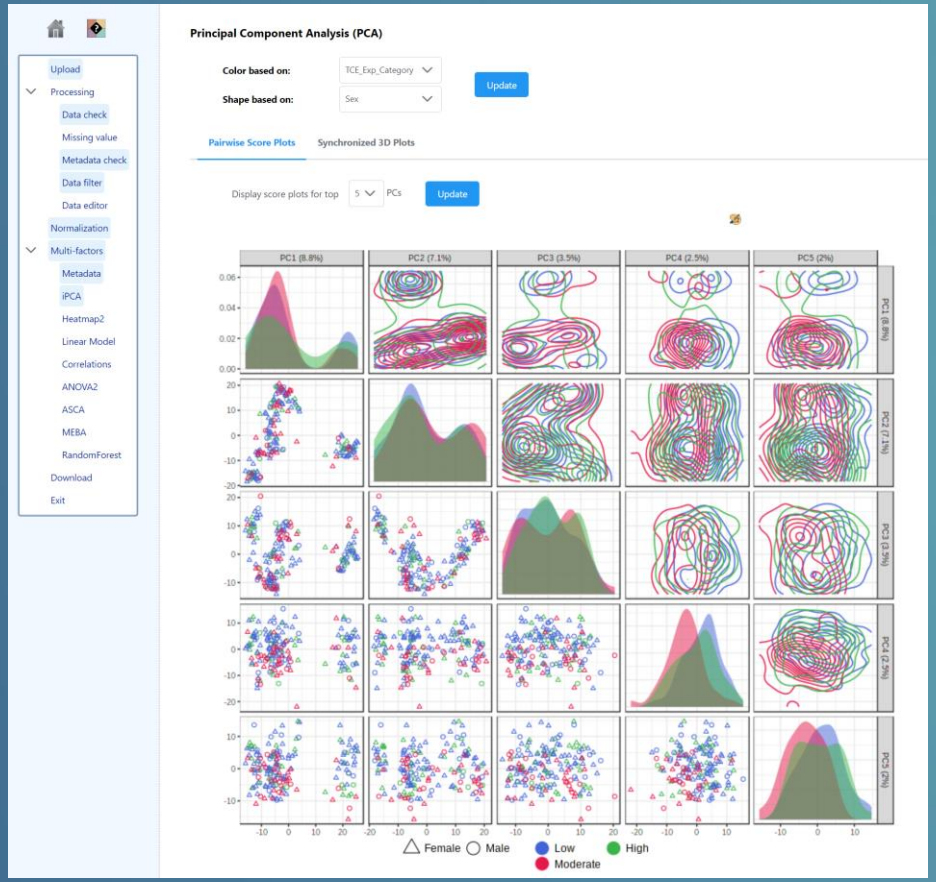
Heatmap of metadata

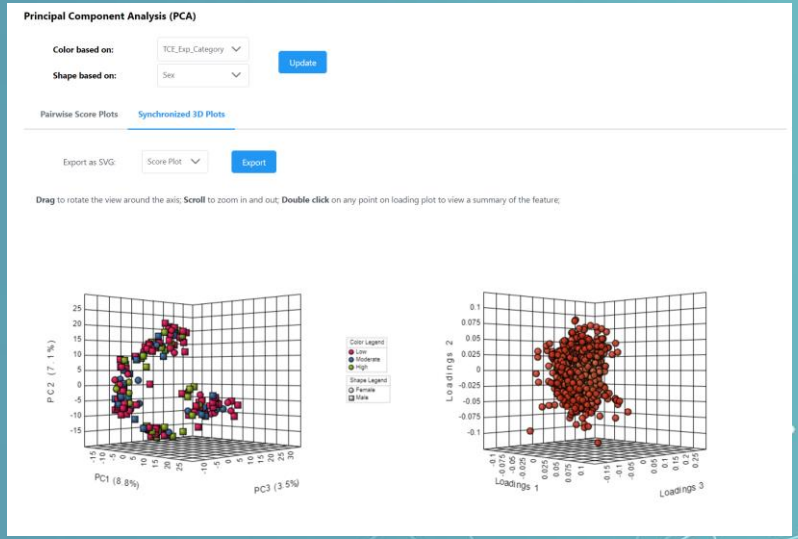


Change correlation method and Click '**Update**'

Correlation analysis of covariates

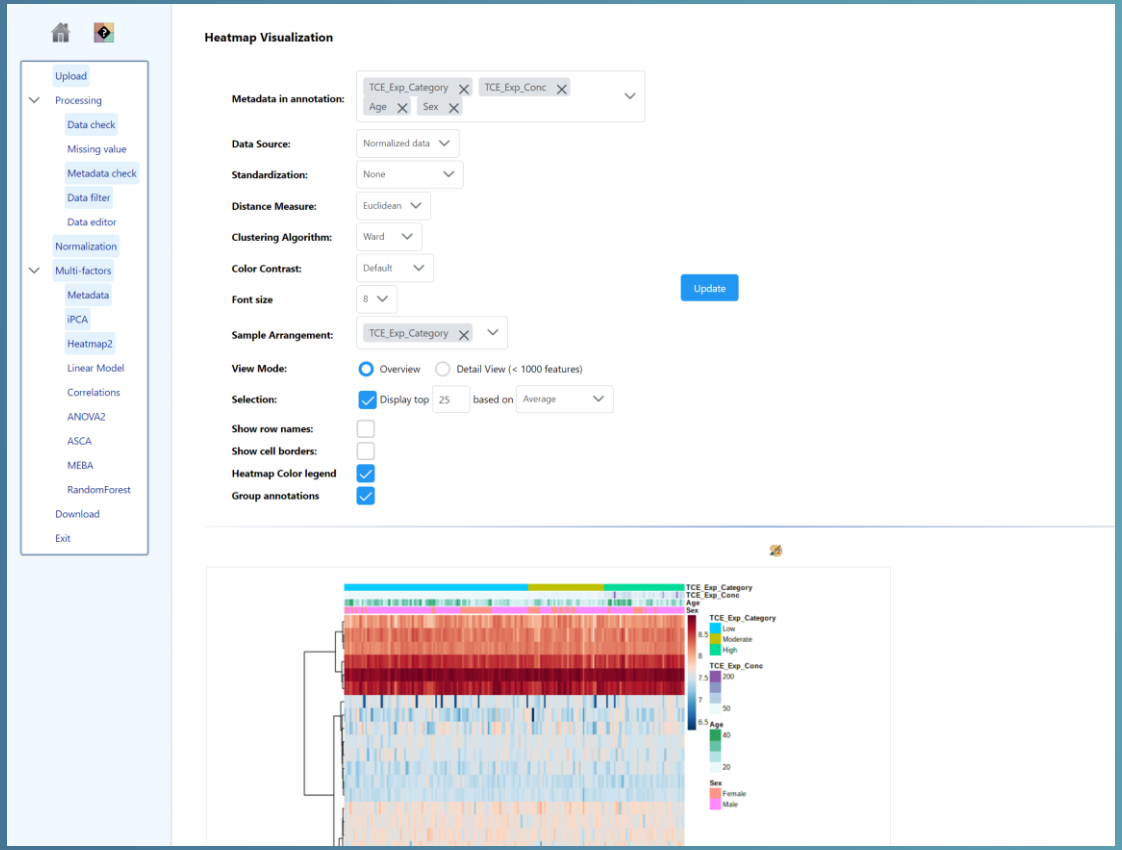
7.7 Interactive PCA





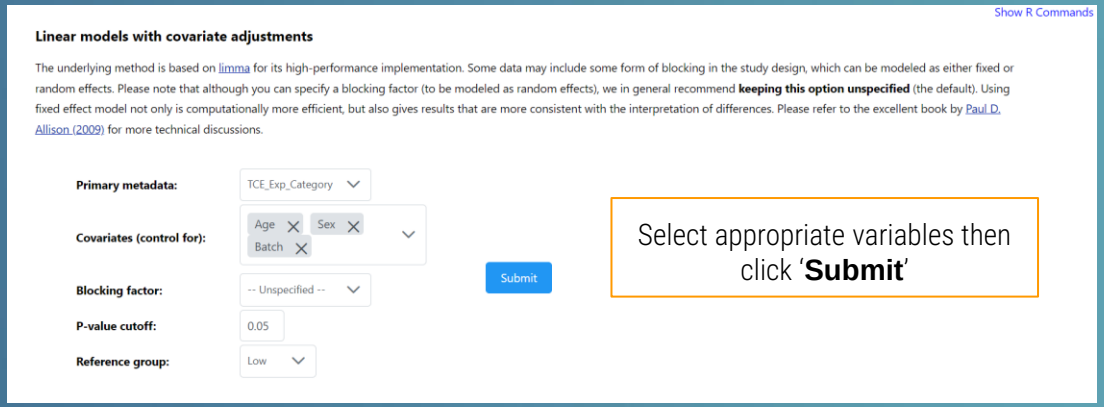
TIP: Principal Component Analysis (PCA) on the data with multiple factors. For the score plot, the PCA are shown as scatter plot and contour plot. The scatter plot shows the details of all samples, while contour plot shows the distribution center of all samples. The density plot on the diagonal display the density of different groups at the corresponding PC based on the primary factor. A pair of dynamic 3D PCA can be explored and rotated easily.

7.8 Hierarchical Clustering and Heatmap Visualization



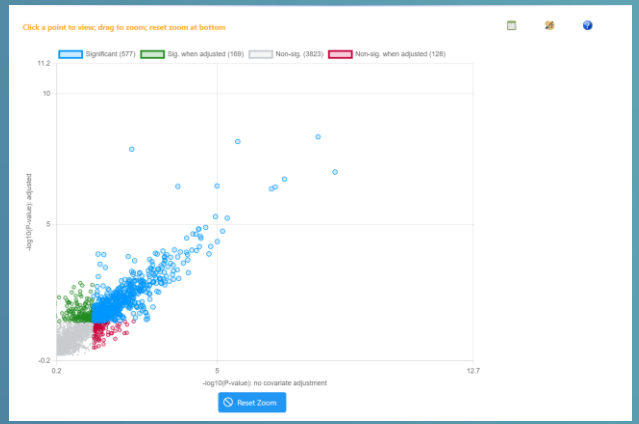
TIP: Clustering of the data can be performed with different methods or based on different factors. Edit the parameters and Click '**Update**' to view the heatmap

7.9 Covariate Adjustment with Linear Models



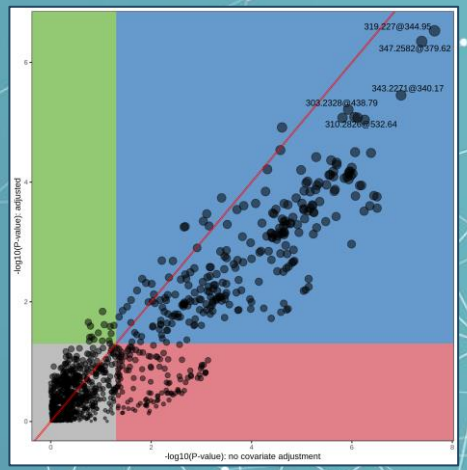
Select appropriate variables then click '**Submit**'

TIP: Find metabolites are significantly associated with your variable of interest ("Primary metadata"), after adjusting for covariates ("Include metadata"). After selecting the appropriate variables, click "**Submit**".

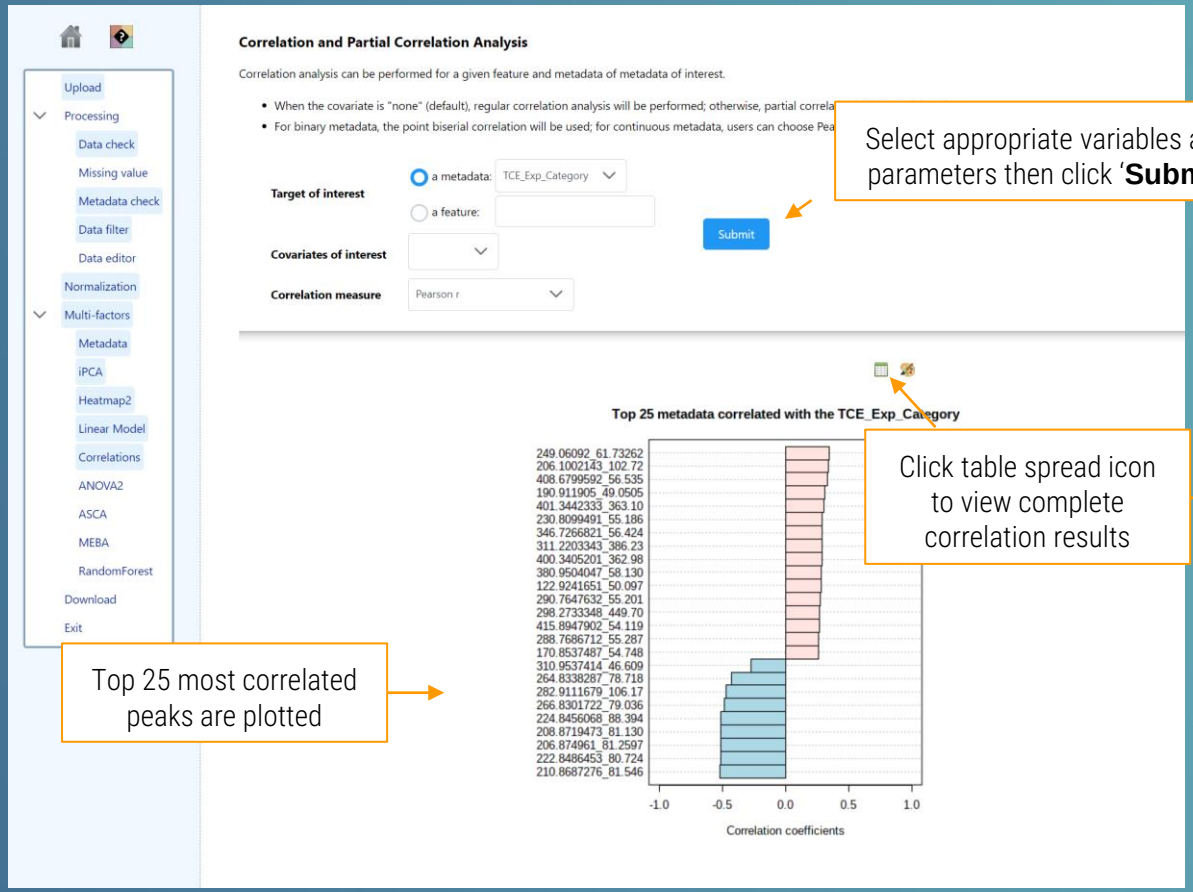


TIP: The results plot compares p-values for each metabolite both before (x-axis) and after (y-axis) covariate adjustment. The green section shows features significant only after adjustment, red are significant only before adjustment, and blue are significant in both cases.

Click the paint icon to generate a high resolution results figure



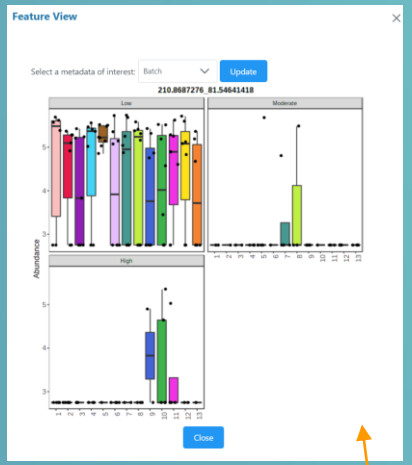
7.10 Correlation Analysis

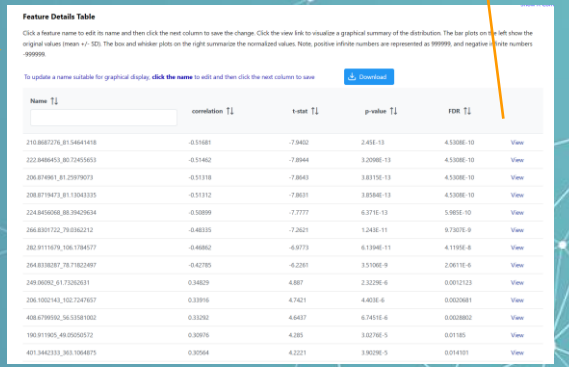


Select appropriate variables and parameters then click '**Submit**'

Click table spread icon to view complete correlation results

Top 25 most correlated peaks are plotted





Click '**View**' for comparison of abundance

7.11 Two-way ANOVA (ANOVA2)

Type of ANOVA:

- For two-factor independent samples: Two-way (between subjects) ANOVA
  - Type I and Type III ANOVA are supported, the main difference is whether the means are weighted (type I) or unweighted (type III).
  - For unbalanced data, there may be signficant difference between type I and type III.
- For time-series only: One-way repeated measures (within subjects) ANOVA
  - The data must have a balanced design (equal sample sizes)
- For time-series + one experimental factor: Two-way repeated measures (within subjects) ANOVA.
  - By default, the ANOVA model will consider the interaction (**Time * Treatment**).
  - If your data does not have sufficient residual degrees of freedom to estimate the interactions, you may want to leave out the interaction term (**Time + Treatment**)

Select two metadata: TCE_Exp_Category × Sex ×

Choose ANOVA Type: Type I (for two-way independent ANOVA)

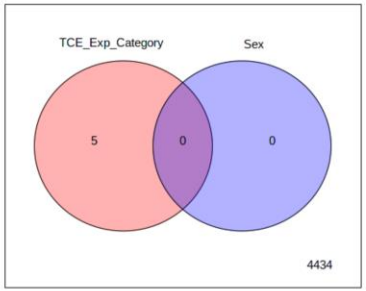
Consider interactions: ☐ (for two-way repeated ANOVA)

Adjusted p-value cutoff: 0.15

Multiple testing correction: False discovery rate

Submit

Two-way ANOVA (between subjects)



Select appropriate variables and parameters then click '**Submit**' to update the Venn Diagram below. Click table spread icon to view more details.

Feature Details Table

Click a feature name to edit its name and then click the next column to save the change. Click the view link to visualize a graphical summary of the distribution. The bar plots on the left show the original values (mean +/- SD). The box and whisker plots on the right summarize the normalized values. Note, positive infinite numbers are represented as 999999, and negative infinite numbers -999999.

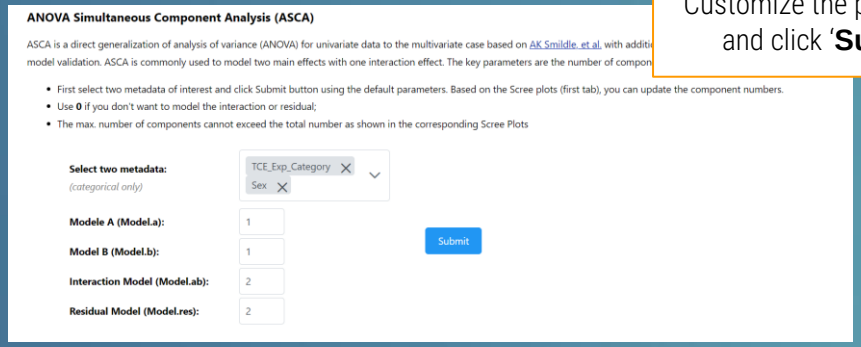
To update a name suitable for graphical display, **click the name** to edit and then click the next column to save

Download

Name	TCE_Exp_Category(F.val)	TCE_Exp_Category(raw.p)	TCE_Exp_Category(adj.p)	Sex(F.val)	Sex(raw.p)	Sex(adj.p)	
206.1002143_102.7247657	11.111	2.9278E-5	0.12997	9.3933E-4	0.97559	0.99741	View
542.9252101_47.92005778	10.221	6.4518E-5	0.13762	0.17591	0.67544	0.98551	View
249.06092_61.73262631	9.806	9.3499E-5	0.13762	0.40834	0.52368	0.9823	View
155.0907026_252.6532109	9.3677	1.3861E-4	0.13762	4.5373	0.03461	0.84297	View
400.3405201_362.9808888	9.2435	1.5501E-4	0.13762	0.15893	0.69065	0.98551	View

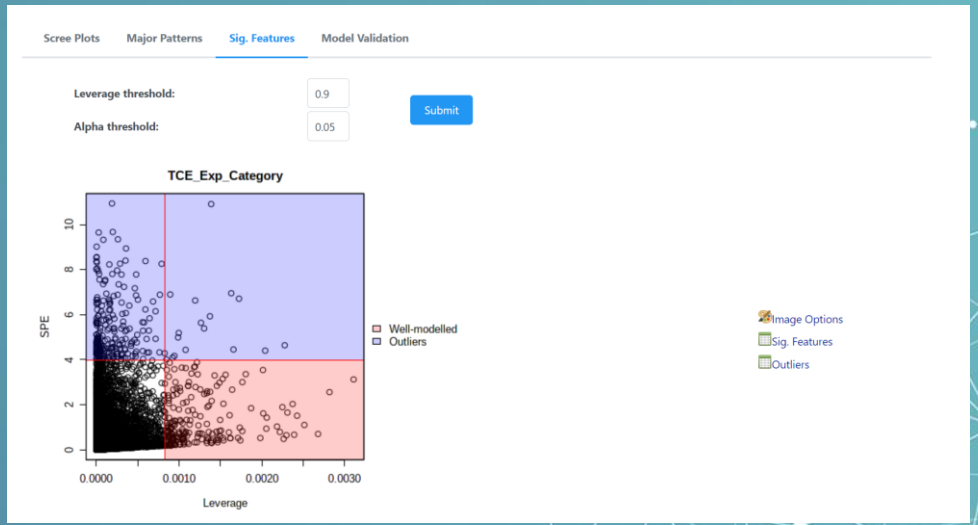
« ‹ 1 › » 20

7.12 ANOVA Simultaneous Component Analysis (ASCA)



Customize the parameters and click '**Submit**'.

View results from different tabs below the parameters panel.



TIP: To fully understand the mechanism of ASCA and interpret the results, please refer to the FAQs section of MetaboAnalyst.

7.15 *Multivariate empirical Bayes (MEBA) time-series analysis*

Time-course profiles:

The table below shows all the features ranked by the corresponding statistics. You can click any feature name to view its corresponding time-course profile.

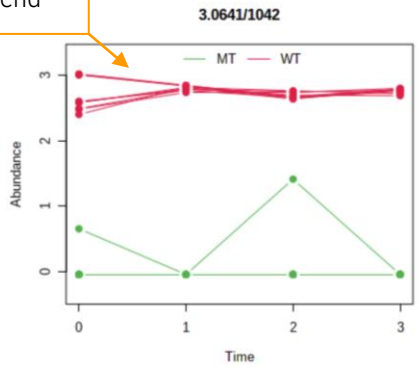
- For time-course only data, we are interested in features which **change** over time.
- For time-course + one experimental factor data, we are interested in features which **respond differently in different experimental conditions** over time.
- For time-course + one experimental factor data, the statistics will be Hotelling-T2 if there are only 2 biological conditions or MB-Statistics if there are more than 2 biological conditions.

Specify metadata: Phenotype ✖ Time ✖ ▾ Update

Customize the metadata included and click '**Update**'.

Name	Hotelling-T2
3.0641/1042	37533.7531
4.0289/873	28526.24962
1.944/1044	23123.73006
3.4428/1019	13978.74526
2.7865/1027	13056.14885
3.108/897	10969.97689
3.0263/811	9867.84225
1.8763/813	9816.35836
3.2347/1012	8767.92088
1.8927/853	7619.41511
1.6114/265	7231.37491
2.7863/1062	6396.38121
4.2231/1096	5607.72909
3.4112/898	5554.81068
4.0271/805	5270.14221
3.0321/869	5266.47452
2.6882/827	5012.98133
3.4422/926	4569.28863
3.1578/841	4288.4332

Select/Click feature name to view the time-series trend



7.14 Supervised Classification with Random Forest

Random Forest

Classification | Var. Importance | Outlier Detection

Leave defaults and click '**Update**'

The model does an o.k. job classifying samples as either "Low" or "Moderate / High".

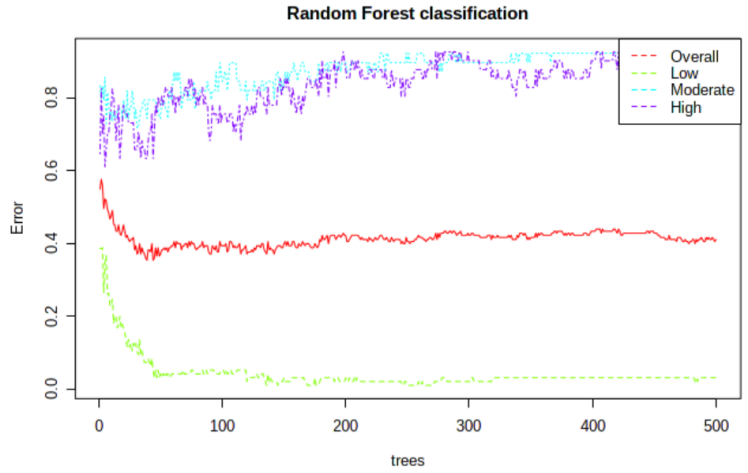
Primary metadata: TCE_Exp_Category

Choose metadata for predictors: Sex

Update

Randomness: On

TIP: To build a predictive model with only metabolites features, leave the default settings. If you'd like to include certain covariates in the model, add them using the "Choose metadata for predictors" dropdown.

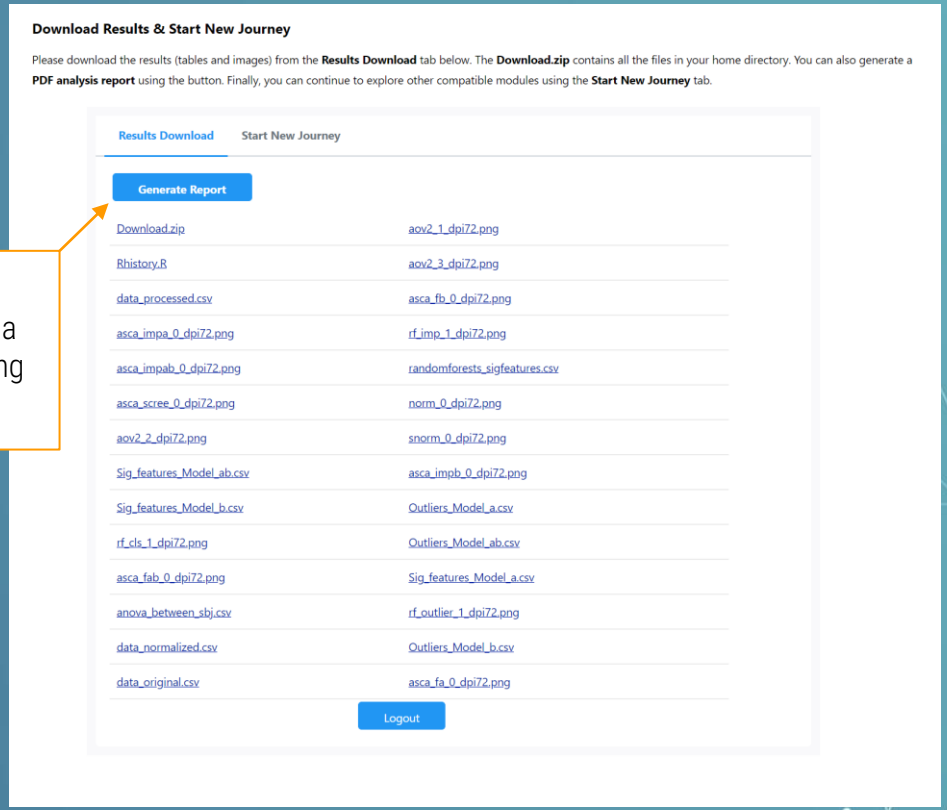


The OOB error is 0.411

	Low	Moderate	High	class.error
Low	92	1	2.0	0.0316
Moderate	33	5	1.0	0.872
High	33	2	6.0	0.854

Look at the results overview

6.4 Result Downloading



Click the "**Generate Report**" to download a pdf report summarizing your analysis.

Thanks

If you have any questions please read through the FAQs or contact us at Zhiqiang.pang[at]xialab.ca or jeff.xia[at]xialab.ca