

## User guide for ExNormalizeMets:

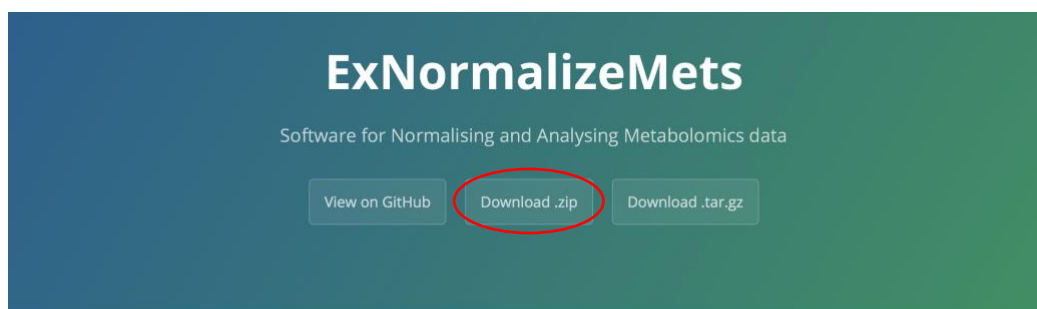
For this guide, a windows machine is being used. No prior software other than Microsoft Excel and a web browser are required for installation. The guide will explain how to install R to be able to use the NormalizeMets R package through its ExNormalizeMets Excel interface (no direct interaction with R or command line functions are needed).

A detailed guide with illustrations for using the software is also provided, showing its main functionalities in a 'walk through' guide of an example project.

For support contact: [olshansky.g@unimelb.edu.au](mailto:olshansky.g@unimelb.edu.au)

### Getting the files ready:

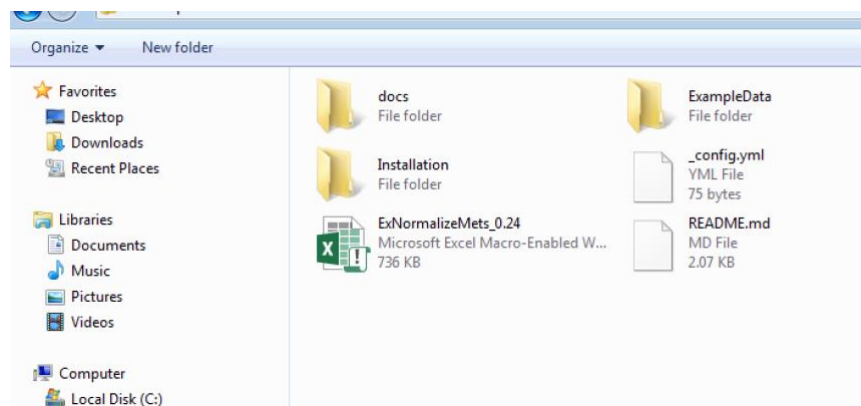
To get the required files, go to <https://metabolomicstats.github.io/ExNormalizeMets/> and download the latest version of the *ExNormalizeMet*:



### ExNormalizeMets

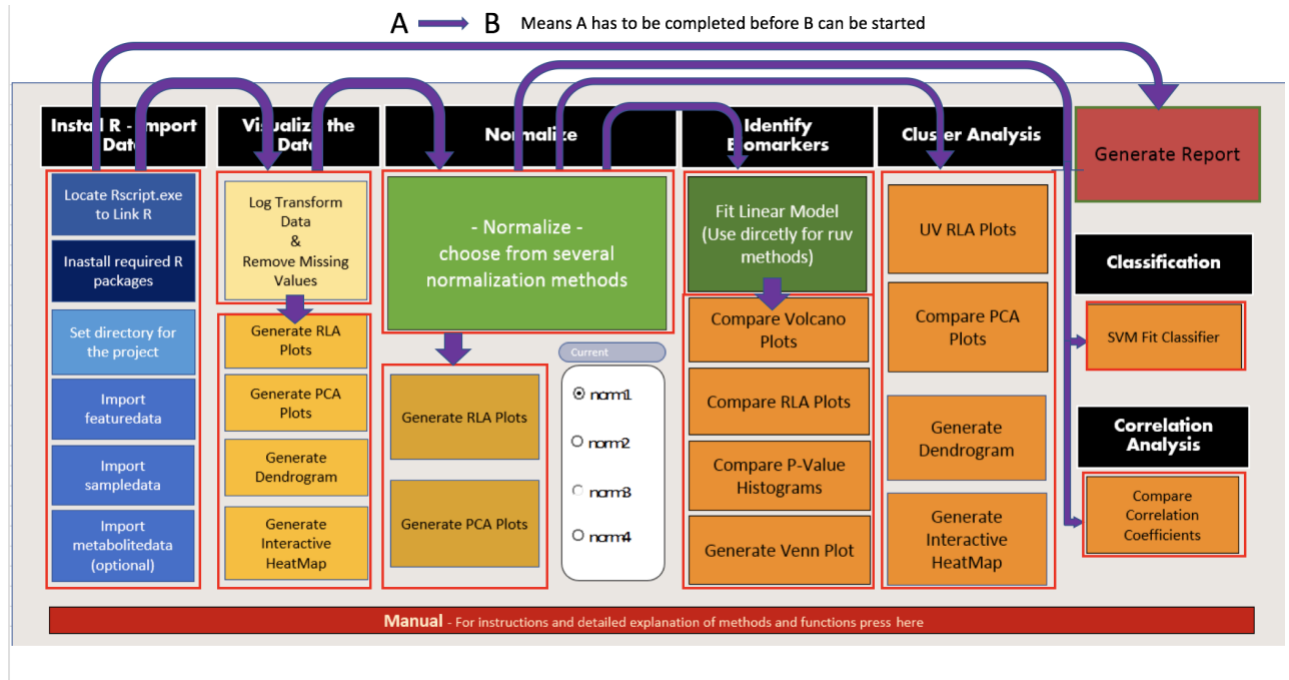
They are all the files needed for installing R and loading NormalizeMets through excel.

Once downloaded, open the folder:



Click on ExNormalizeMets\_0.24 to open the excel worksheet but otherwise don't modify the files in this folder (the downloaded folder can be copied to a different location on your computer for convenient access although it is only needed for first use).

## General Workflow:



## Installing R:

1. Install R from CRAN (The Comprehensive R Archive Network) by going to the following page:

<https://cran.r-project.org>



CRAN  
Mirrors  
What's new?  
Task Views  
Search

About R  
R Homepage  
The R Journal

Software  
R Sources  
R Binaries  
Packages  
Other

Documentation  
Manuals  
FAQs  
Contributed

### The Comprehensive R Archive Network

#### Download and Install R

Precompiled binary distributions of the base system and contributed packages, **Windows and Mac** users most likely want one of these versions of R:

- [Download R for Linux](#)
- [Download R for \(Mac\) OS X](#)
- [Download R for Windows](#)

R is part of many Linux distributions, you should check with your Linux package management system in addition to the link above.

#### Source Code for all Platforms

Windows and Mac users most likely want to download the precompiled binaries listed in the upper box, not the source code. The sources have to be compiled before you can use them. If you do not know what this means, you probably do not want to do it!

- The latest release (2017-11-30, Kite-Eating Tree) [R-3.4.3.tar.gz](#), read [what's new](#) in the latest version.
- Sources of [R alpha and beta releases](#) (daily snapshots, created only in time periods before a planned release).
- Daily snapshots of current patched and development versions are [available here](#). Please read about [new features and bug fixes](#) before filing corresponding feature requests or bug reports.
- Source code of older versions of R is [available here](#).
- Contributed extension [packages](#)

#### Questions About R

- If you have questions about R like how to download and install the software, or what the license terms are, please read our [answers to frequently asked questions](#) before you send an email.

#### What are R and CRAN?

R is 'GNU!'S' a freely available language and environment for statistical computing and graphics which provides a wide variety of statistical and



## R for Windows

Subdirectories:

[base](#)

Binaries for base distribution. This is what you want to [install R for the first time](#).

[contrib](#)

Binaries of contributed CRAN packages (for R  $\geq$  2.13.x; managed by Uwe Ligges). There is also information on [third party software](#) available for CRAN Windows services and corresponding environment and make variables.

[old contrib](#)

Binaries of contributed CRAN packages for outdated versions of R (for R  $<$  2.13.x; managed by Uwe Ligges).

[tools](#)

Tools to build R and R packages. This is what you want to build your own packages on Windows, or to build



## R-3.4.3 for Windows (32/64 bit)

[Download R 3.4.3 for Windows](#) (62 megabytes, 32/64 bit)

[Installation and other instructions](#)

[New features in this version](#)

If you want to double-check that the package you have downloaded matches the package distributed by CRAN, you can compare the [md5sum](#) of the .exe to the [fingerprint](#) on the master server. You will need a version of md5sum for windows: both [graphical](#) and [command line versions](#) are available.

### Frequently asked questions

- [Does R run under my version of Windows?](#)
- [How do I update packages in my previous version of R?](#)
- [Should I run 32-bit or 64-bit R?](#)

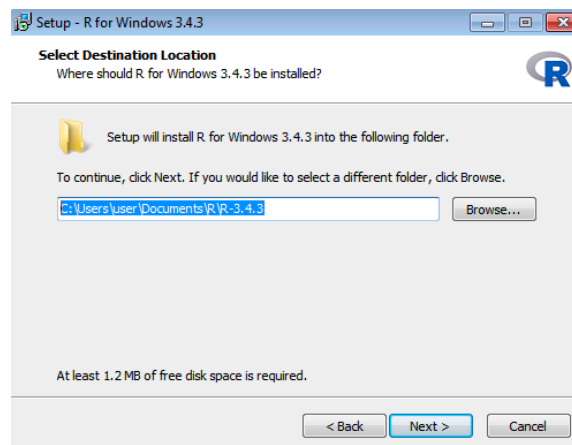
Please see the [R FAQ](#) for general information about R and the [R Windows FAQ](#) for Windows-specific information.

### Other builds

- Patches to this release are incorporated in the [r-patched snapshot build](#).
- A build of the development version (which will eventually become the next major release of R) is available in the [r-devel snapshot build](#).
- [Previous releases](#)

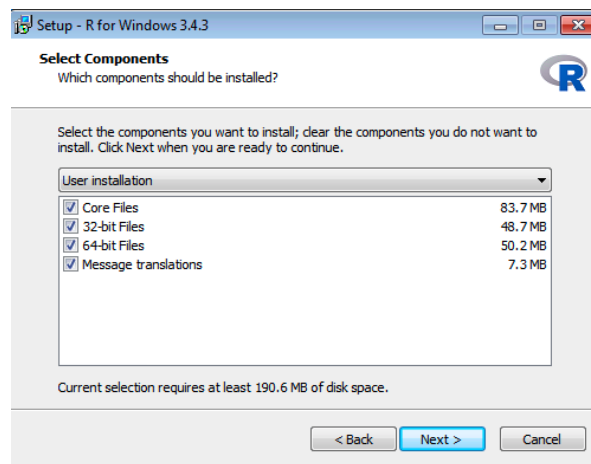
Note to webmasters: A stable link which will redirect to the current Windows binary release is [<CRAN MIRROR>/bin/windows/base/release.htm](#).

2. Once the installation is complete, run the downloaded file R-3.4.3-win.exe, follow the installation instructions and choose the location where R should be installed.

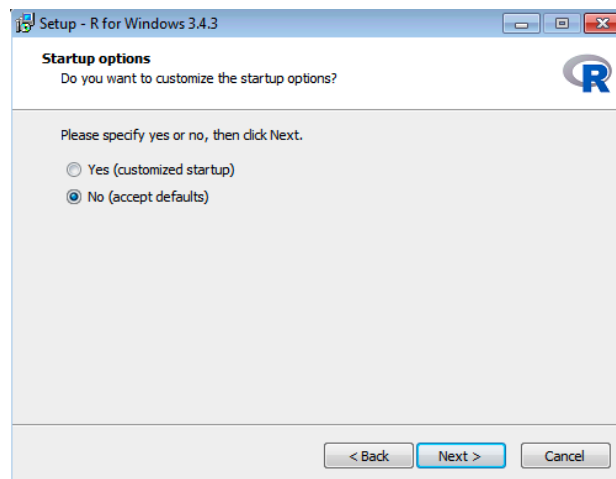


Click *Next* for default location (Recommended) or choose location manually. Make sure you **remember where R is installed** as you will need to locate this folder later to link R to Excel.

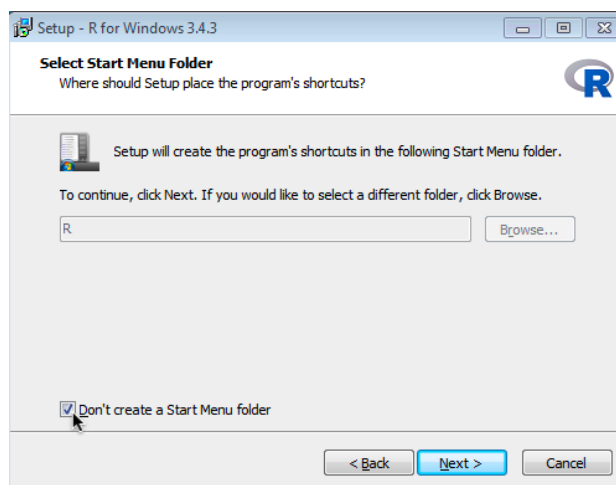
3. Click *Next* to install with the different required settings:

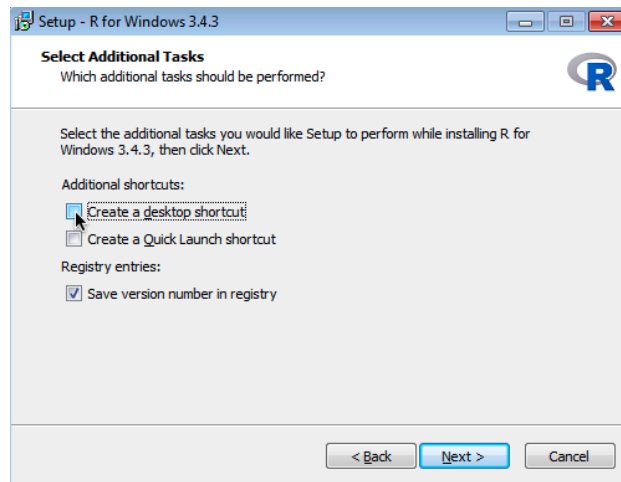


4. Click *Next* to install with default settings:

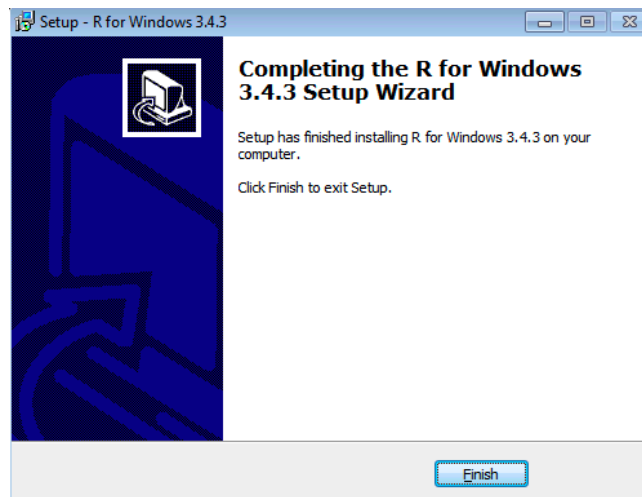


5. In the next screen check the *Don't create a Start Menu folder* icon if you don't intend to use R by itself and click next. Also uncheck the *Create a desktop shortcut* in the next screen:



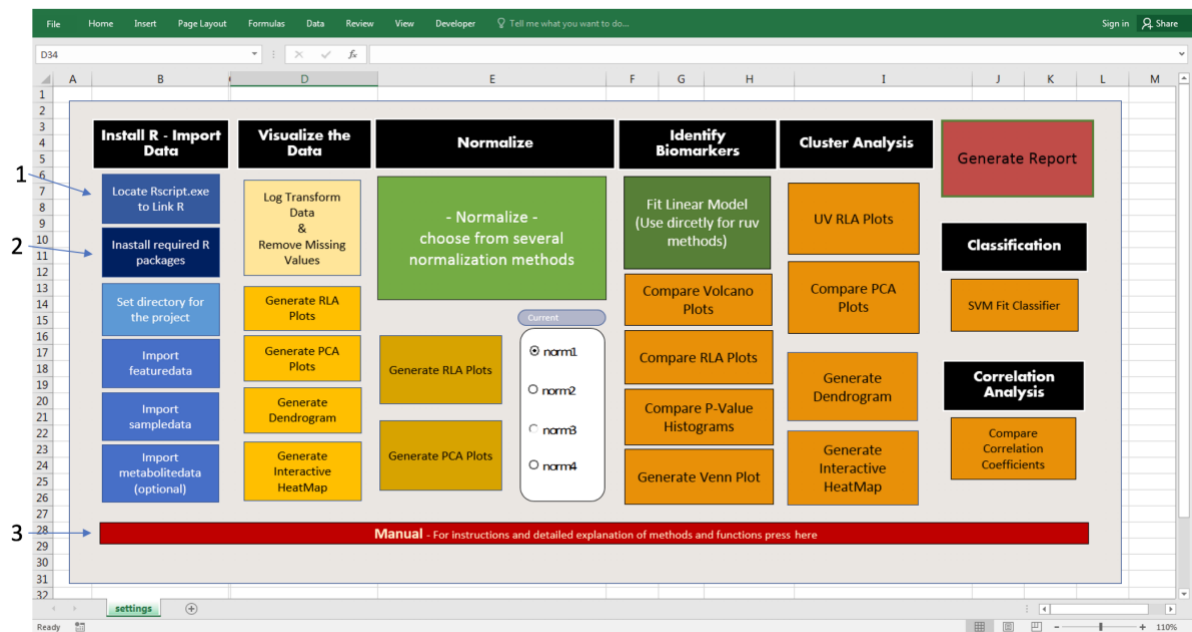


6. Wait until installation is done, the following screen should appear



### **Linking R to excel and installing the required packages (first use only):**

After installing R open the excel file ExNormalizeMets.xlsm, this will open the Excel interface onto the settings sheet:

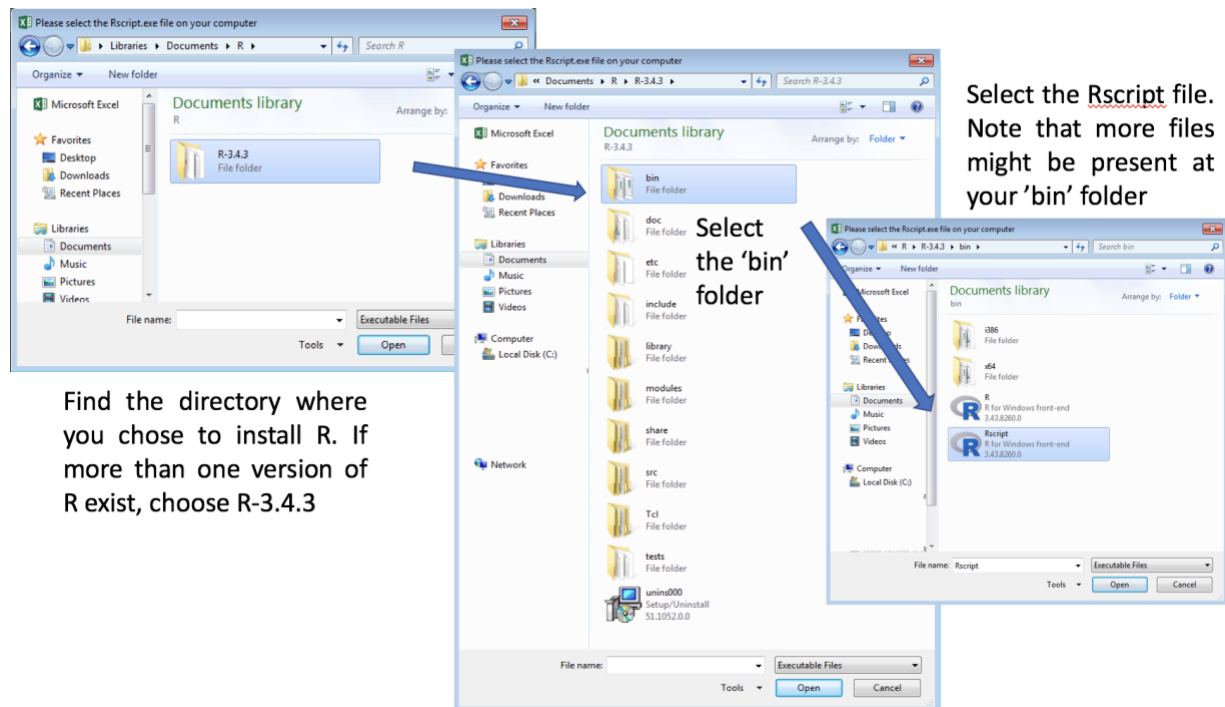


The *settings* sheet is your ‘**Control sheet**’, any function you want to run, from importing data, Normalizing, viewing results and opening the manual can be done from here.

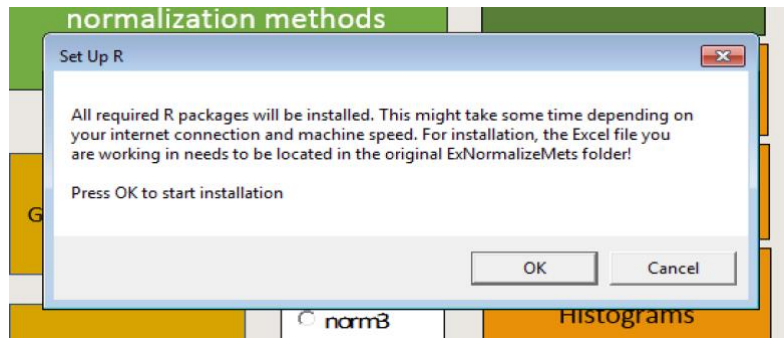
This manual can be accessed at any stage by clicking on (3) but more on this later. For now, first locate the Rscripts.exe file so that excel will know how to run R commands it generates.

## Locating Rscript:

Press (1. ‘*Locate Rscript.exe to Link R*’) to locate the Rscript file in the window that opens up. Make sure to select the *Rscript* file in the *bin* folder of your R installation:

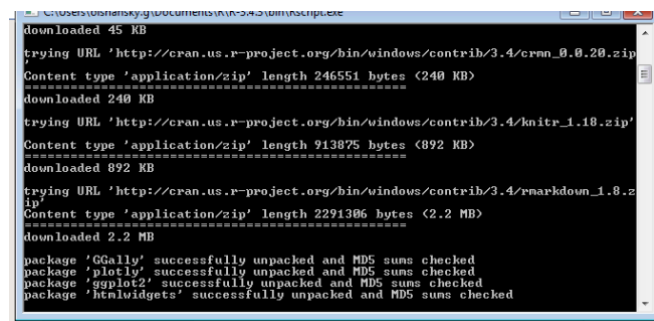


After selecting the file, to install all the required R packages and set up the needed dependencies, press (2. ‘*Install required R packages*’).



Pressing ok will start the installation, this might take a few minutes if you are using R for the first time as many of the base packages will need to be installed.

The installation window looks like this:



When the installation is done, a window with the message *Done!* will appear.

Now that the installation is complete, NormalizeMets is ready for use!

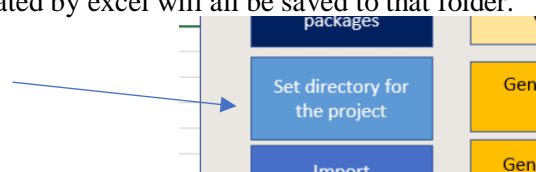
## Using NormalizeMets

Data used in the following examples is provided with NormalizeMets (alldata\_eg in R), it is located in the *ExampleData* folder in your downloaded *ExNormalizeMetsSetup* file. Future references in this guide refer to this data by default.

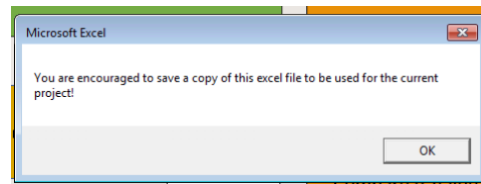
An example excel document containing the data used for the tutorial with all settings identical to those in the tutorial is provided (MyFirstNormalizeMetsProject\_example.xlsm), if using this document, make sure to set the Rscript location and working directory for your machine.

### Starting a new project:

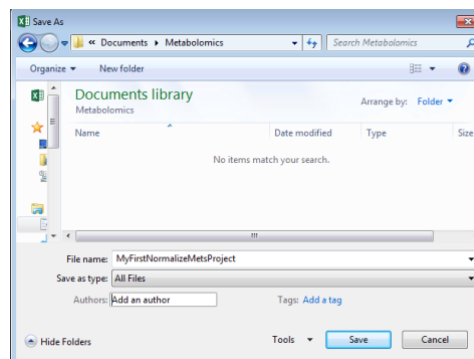
To start a new project, you will need to set up or use an existing directory where the project will ‘live’, data and plots generated by excel will all be saved to that folder.



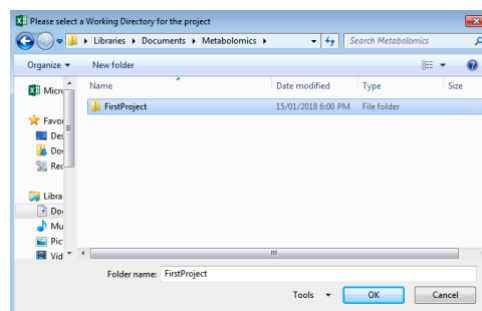
Before selecting the folder, you will be prompted to save a copy of the current version of the excel file, it is recommended to save it with a new name to make sure a 'clean' version always stays in your ExNormalizeMetsSetup folder.



After saving the workbook under the name of your choice:

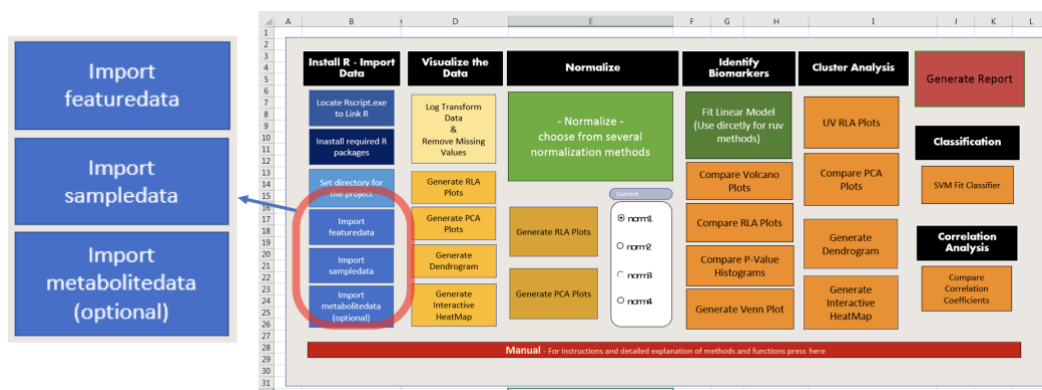


Select the working directory where all files should be generated. We recommend making a new folder for each project.



## Loading data:

To load data for the project, in turn click on the following to load the relevant data:





Loaded data needs to be in .csv format. After clicking on the required file, it will open a new sheet, showing the imported data. Select the setting sheet import more data and get back to the options.

For *featuredata*, set metabolites in columns and samples in rows. Unique sample names should be provided as row names.

	m_1	m_2	m_3	m_4	m_5	m_6	m_7	m_8	m_9	m_10	m_11	m_12
s_1	10485.86719	33220.5625	1112.979492	1408.630648	455.7529297	1100.402344	122.3631592	3855.804688	1637.482422	6194.292969	28793.65625	5200.7
s_2	8960.46875	29995.51563	926.2890625	529.0449219	873.609375	2201.246094	173.0802002	5090.636719	2011.500977	8868.859375	31683.09375	6183.9
s_3	10160.44531	28559.64063	1230.333008	1306.320313	1027.507813	2066.591797	269.7998047	4483.90625	1644.607422	7776.335938	44494.59375	10505.1
s_4	8794.476563	27593.75	901.762207	1800.083008	675.0395508	1675.849609	287.9992676	4949.261719	1508.802734	8405.171875	37030.40625	10726.1
s_5	8956.921875	28161.76563	979.1723633	818.0366211	904.0253906	1245.991211	167.4133301	4302.933594	1460.012695	6976.429688	29579.53125	5421.9
s_6	9092.257813	31685.3125	634.6899414	478.3811015	980.4233398	1259.563477	97.5625	4406.34375	2349.587891	7090.136719	39835.71875	5657.8
s_7	2271.044922	7692.175781	143.1688232	109.8399048	730.6171875	1387.791992	149.0966797	1543.40825	961.7314453	3356.685547	20212.70313	4483.1
s_8	7850.402344	26462.79688	822.5991211	1341.21582	583.5830078	2164.126953	227.7696533	5030.6875	1604.274414	6241.457031	30873.14063	8130.1
s_9	10969.0625	21605.14063	408.0004883	1499.479492	452.2375488	916.5283203	234.4975586	4657.253906	1847.6875	6606.15625	34395.96875	7142.3
s_10	1743.932617	7647.644531	218.4199219	127.4294434	408.2912598	999.5717773	74.82696533	1510.011719	895.9853516	2407.103516	18256.14063	2358.1
s_11	1284.412109	4819.996094	61.62762451	15.51902008	394.8364258	1112.143555	28.35443115	1163.742188	628.7016602	2646.070313	10544.94531	1728.1
s_12	2272.923828	6813.679688	243.9420166	280.7983398	434.3505859	1167.960938	86.25360107	1665.297852	1009.748047	3927.849609	21808.45313	6546.9
s_13	9838.0625	28486.11875	577.8486328	1302.723633	350.9999609	1200.836719	134.2883301	4406.34375	1777.506836	8467.085938	30871.17188	6432.0
s_14	6795.733906	24264.64063	745.7470709	405.0424805	990.1313477	1794.898438	119.0358276	4313.539063	1480.686523	6845.367188	32930.875	8029.3
s_15	1836.59375	6009.238281	42.959473	37.65142822	491.6550293	1753.573242	70.06390381	1526.198242	941.4536133	3346.365234	20823.03125	2188.4
s_16	1699.289063	6458.503906	96.57415771	200.644165	310.729248	774.0219727	183.9802246	1690.515625	900.3242188	3143.683594	14266.88281	1877.8
s_17	1497.520508	5609.605469	77.26855469	169.138916	226.6845703	567.8823242	79.01464844	1362.202148	830.9599609	2024.548828	9131.3125	1730.2
s_18	9033.570313	21372.0625	1250.067383	1564.307617	892.3813477	1589.588867	252.1553955	4251.226563	1510.931641	6424.839844	38623.5625	7712.2

*sampeldata* should have sample information matching featuredata (samples in rows).

	batch	gender	Age	bmi
s_1	Batch 2 code_1	1	58.7	22.2
s_2	Batch 2 code_0	76.7	23.7	
s_3	Batch 2 code_0	56.2	28.2	
s_4	Batch 2 code_0	77.2	26.2	
s_5	Batch 2 code_0	74.3	26.4	
s_6	Batch 2 code_1	66.8	26.4	
s_7	Batch 1 code_0	65	29	
s_8	Batch 2 code_1	66.5	26.1	
s_9	Batch 2 code_0	70.2	27.3	
s_10	Batch 1 code_1	55	25	
s_11	Batch 1 code_0	51	28	
s_12	Batch 1 code_1	55	29	
s_13	Batch 2 code_0	79.2	30.7	
s_14	Batch 2 code_1	52.7	25.4	
s_15	Batch 1 code_0	42	32	

Optional *metabolitedata* should have metabolite information matching featuredata with metabolite names in rows.

	names	IS_neg_controls	pos_controls	gender
m_1	0	1	0	
m_2	0	1	0	
m_3	0	1	0	
m_4	0	0	0	
m_5	0	0	0	
m_6	0	0	0	
m_7	0	0	0	
m_8	0	1	0	
m_9	0	1	0	
m_10	0	1	0	
m_11	0	0	0	
m_12	0	1	0	
m_13	1	0	0	
m_14	0	0	0	

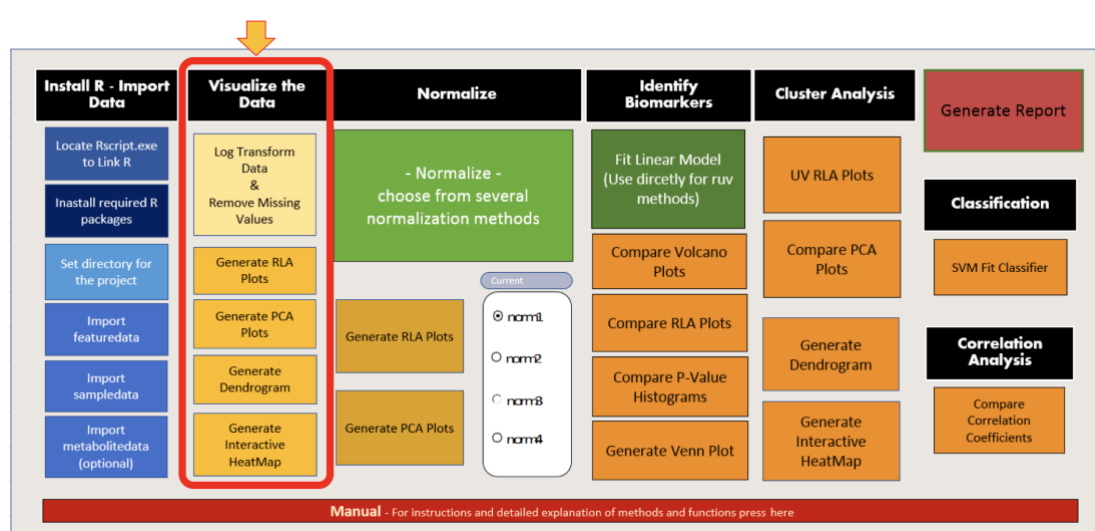
*metabolitedata* can include any metabolite information such as grouping structures, internal standard metabolites, negative control and positive control metabolites.

After the data is loaded, you are ready to proceed to analyse the data!

## NormalizeMets Workflow:

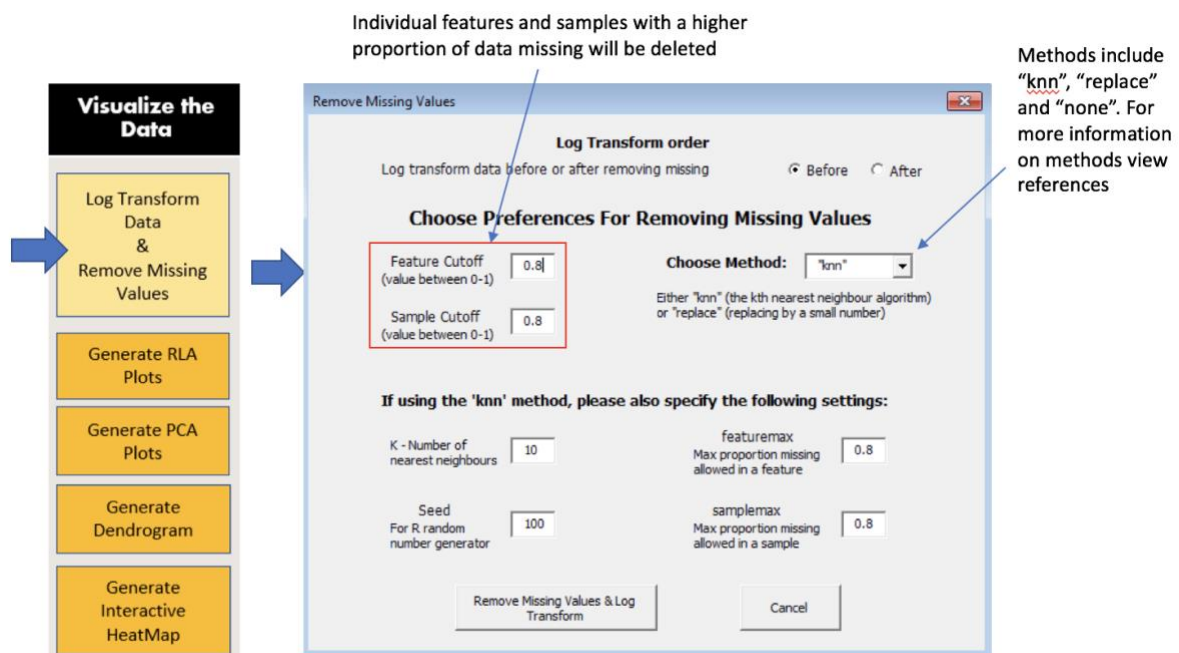
## Visualize the Data

The following section refers to the visualize part:



### Log Transforming the data and removing missing value (mandatory):

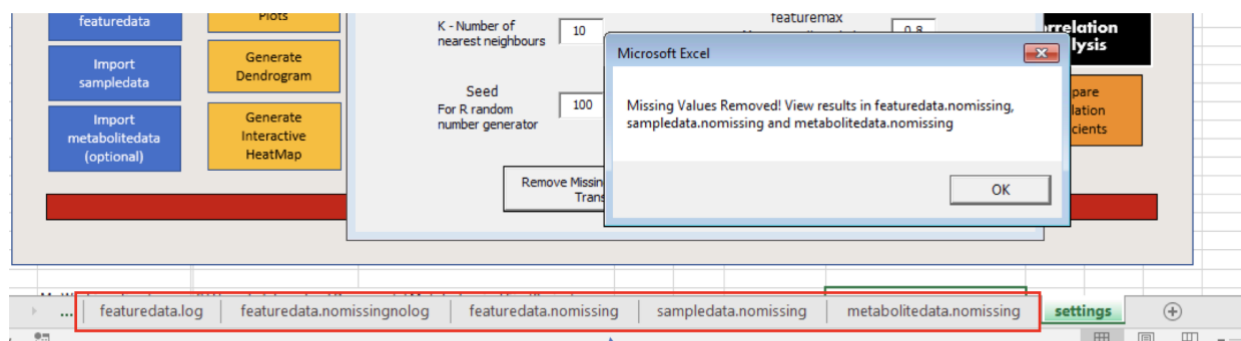
A frequent issue in metabolomics data sets is the occurrence of missing values. It is important to reduce the number of missing values as much as possible by using an effective pre-processing procedure. For example, a secondary peak picking method can be used for LC-MS data to fill in missing peaks which are not detected and aligned.



“knn” – use k nearest neighbours method to replace missing values.

“replace” – replaces missing values by half the minimum value in featuredata.

Clicking ‘Remove Missing Values & Log Transforming’ the following appears:

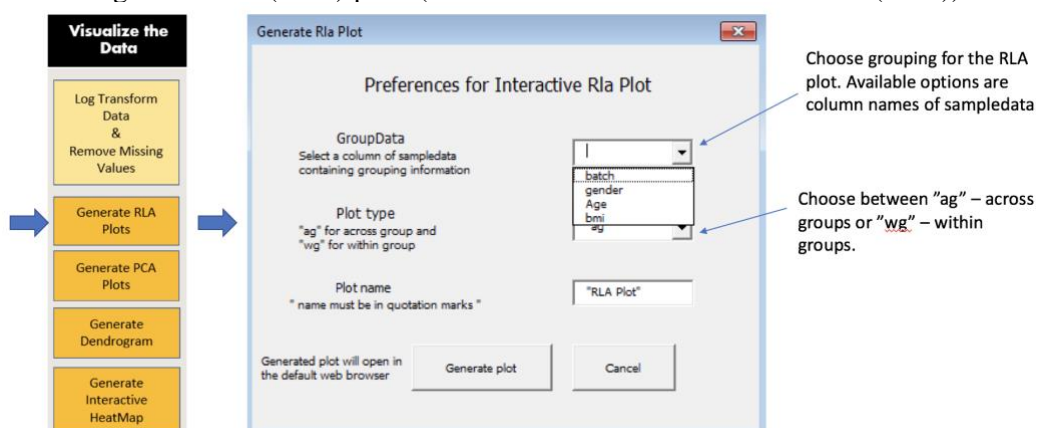


Note the new sheets that appeared, they have respectively the regular log transformed featuredata, the log transformed featuredata with missing values removed, featuredata with missing values removed and without log transformation, sampleddata with rows removed corresponding to featuredata.nomissing, metabolitedata with rows removed corresponding to featuredata.nomissing. Unless you are interested to view or copy any of this data, those sheets are only going to be used for further internal functions.

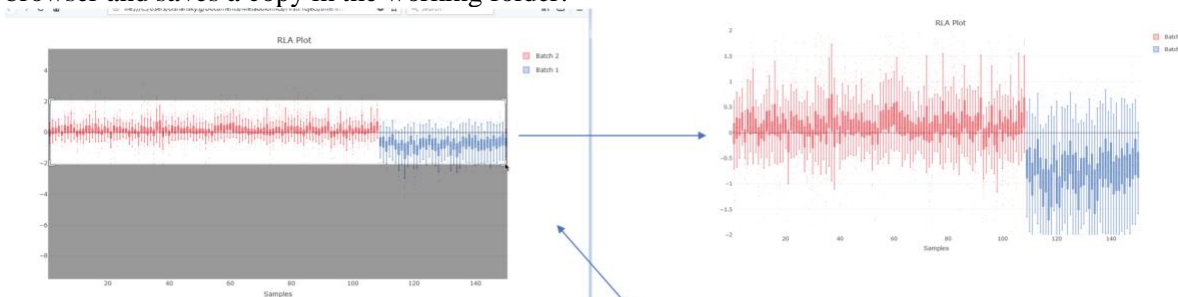
Now the plots in *Visualize the Data* can be generated! The data generated is also going to be used for the Normalization section.

## RLA plots

One way of visualising the log transformed metabolomics data is the use of *across group* or *within group* relative log abundance (RLA) plots (De Livera et al. 2012 De Livera et al. (2015)).

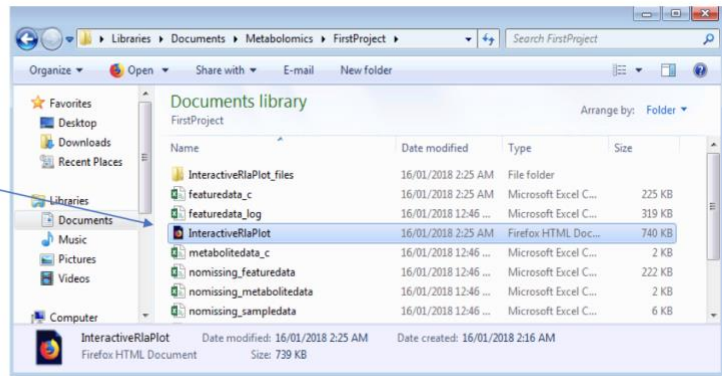


Setting groupdata to *batch* and selecting *Generate plot* opens the interactive plot in the default web browser and saves a copy in the working folder.



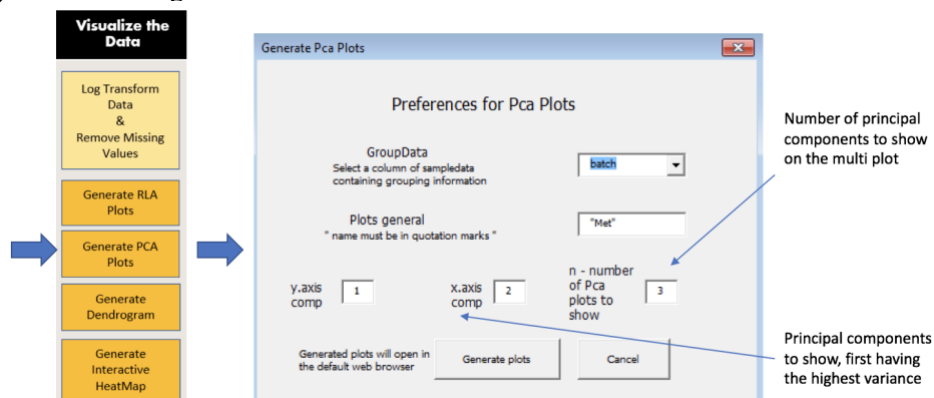
Interactive plot opens in the default browser, to zoom in, simply select the required part

Copy of all plots together with all other generated files are stored in the working directory



## PCA plots

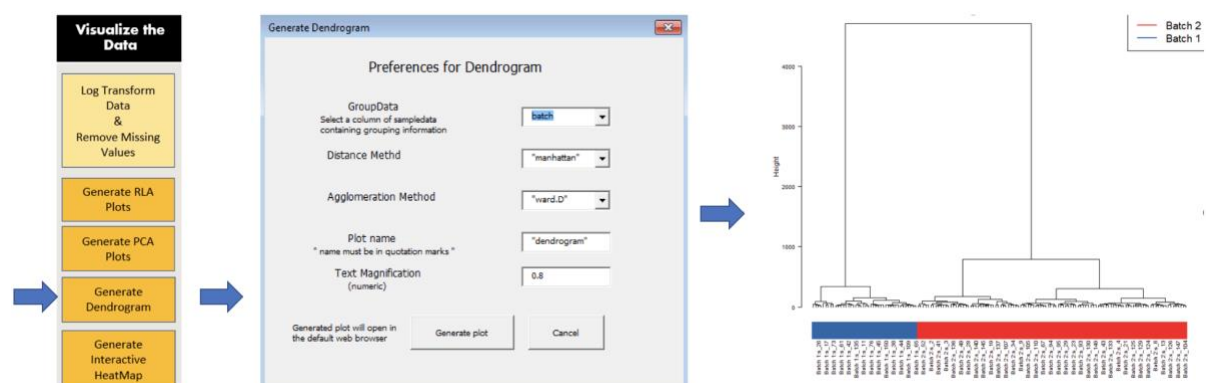
The following function can be used to obtain multiple plots for exploration of the principal components of the *featuredata* matrix: a bar plot indicating the variance explained by each principal component, scores and loading plots with specified axes (interactive and non-interactive), and a pairs plot of the first *n* principal components. These plots are useful in identifying any outlying samples and getting a preliminary understanding of the structure of the data.



All produced plots are stored in the working directory, with interactive plots opened in the browser and static plots located in the new *plots* sheet.

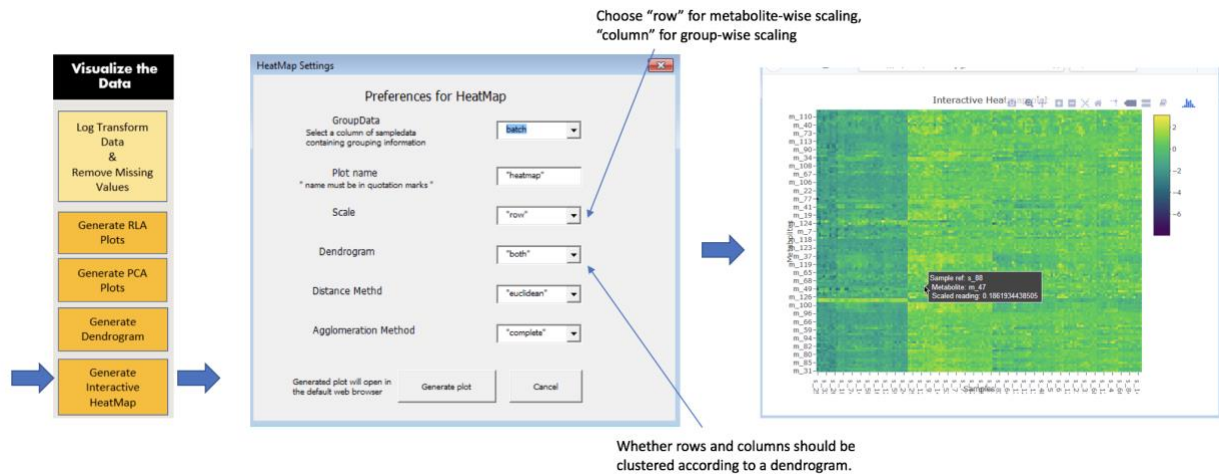
## Dendrogram

Generates a dendrogram to visualise clustering structures in the data, many different methods are available.



## HeatMap

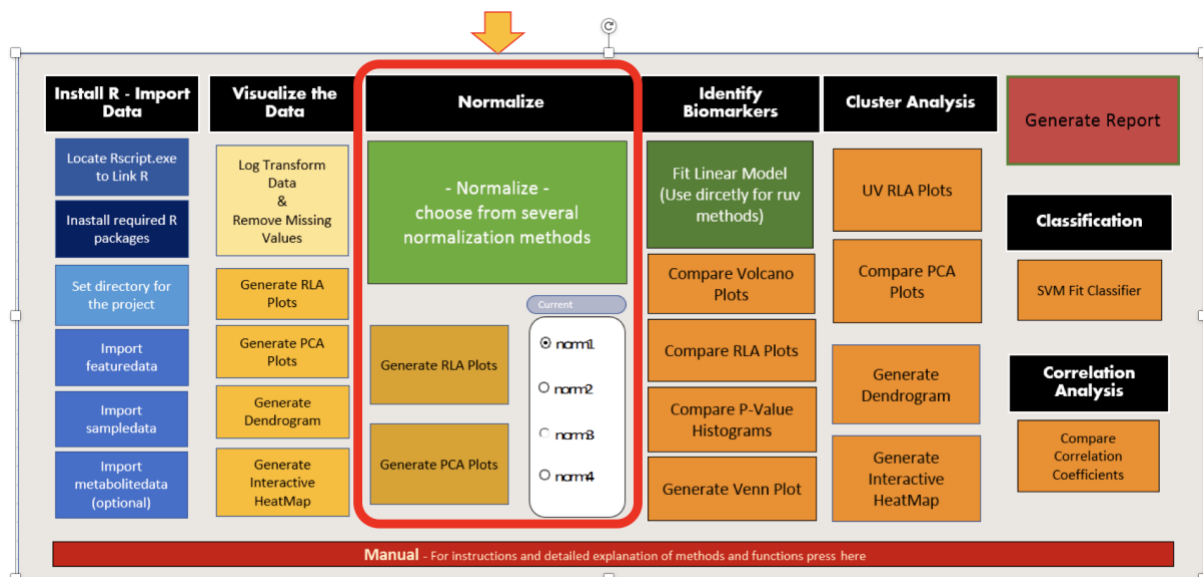
The HeatMap produced can reveal interesting structures in the data.



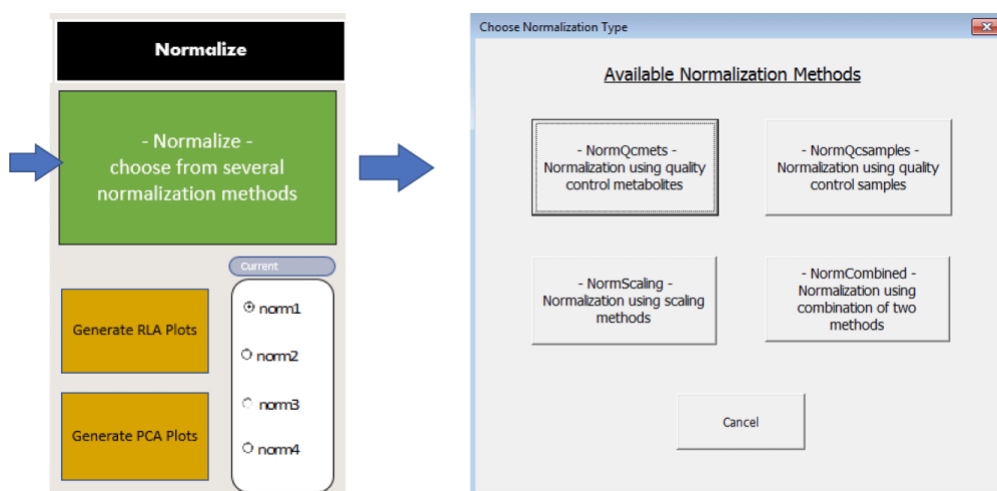
## Normalization

Normalization methods presented in this package are divided into four categories, as those which use (i) internal, external standards and other quality control metabolites (*NormQcmets*) (Sysi-Aho et al. 2007, Redestig et al. (2009), De Livera et al. (2012), De Livera et al. (2015), Gullberg et al. (2004)) (ii) quality control samples (*NormQcsamples*) (Dunn et al. 2011), (iii) scaling methods (*NormScaling*) (Scholz et al. 2004, Wang et al. (2003)), and (iv) combined methods (*NormCombined*) (Kirwan and Broadhurst (2013)).

The normalization methods are accessible in the following section:



Clicking on the Normalize button opens the following menu enabling the choice of different normalization methods.

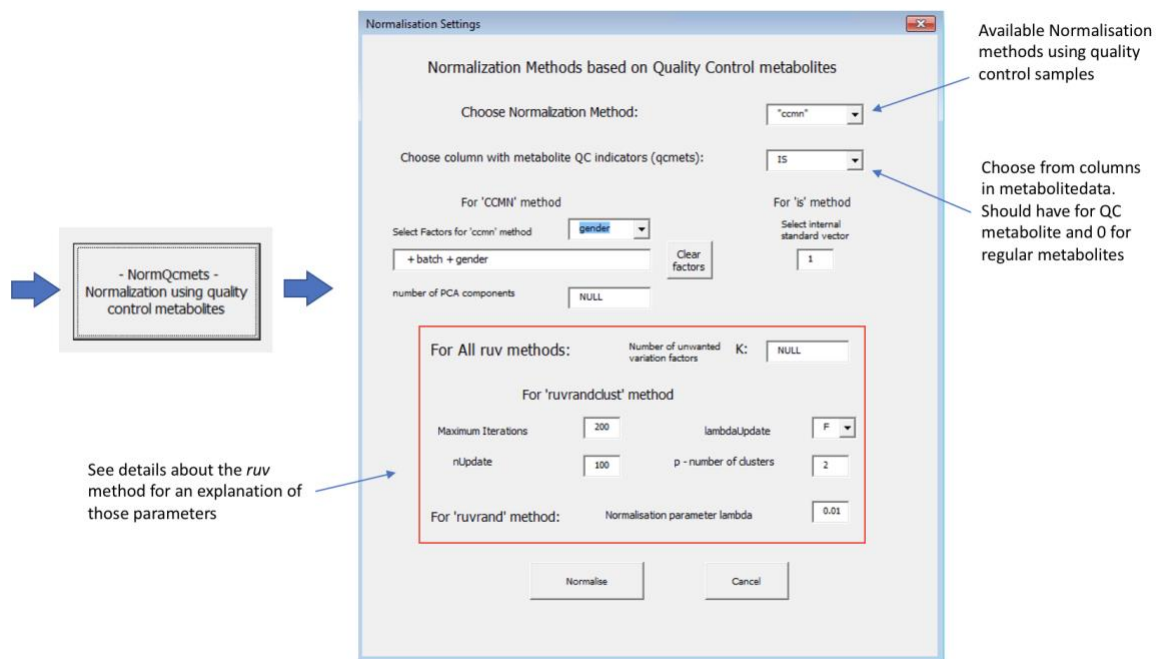


## NormQcmets

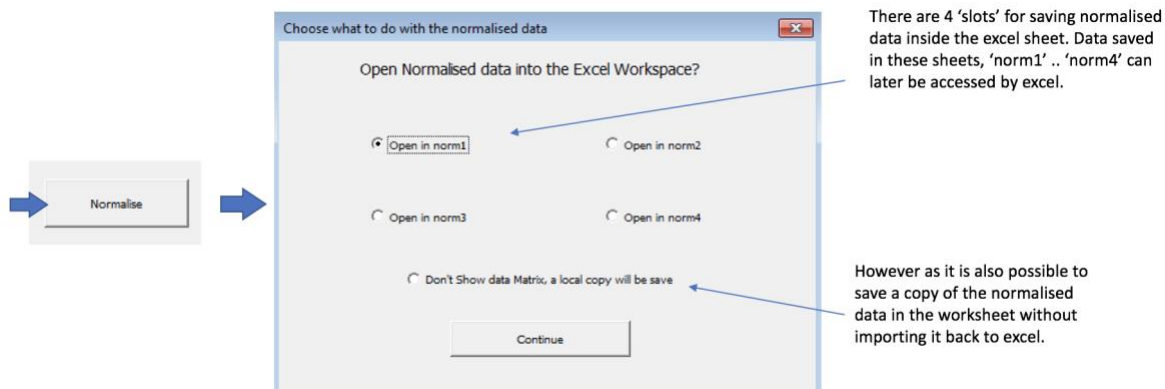
The approaches in *NormQcmets* use internal, external standards and other quality control metabolites. These include the *is* method which uses a single standard (Gullberg et al. 2004), the *ccmn* (cross contribution compensating multiple internal standard) method (Redestig et al. 2009), the *nomis* (normalization using optimal selection of multiple internal standards) method (Sysi-Aho et al. 2007), and the remove unwanted variation methods (J. A. Gagnon-Bartsch, Jacob, and Speed 2014) as applied to metabolomics using “*ruv2*” (De Livera et al. 2012), “*ruvrand*” and “*ruvrandclust*” (De Livera et al. 2015). Note that *ruv2* is an application specific method designed for identifying biomarkers using a linear model that adjusts for the unwanted variation component.

To Normalize:

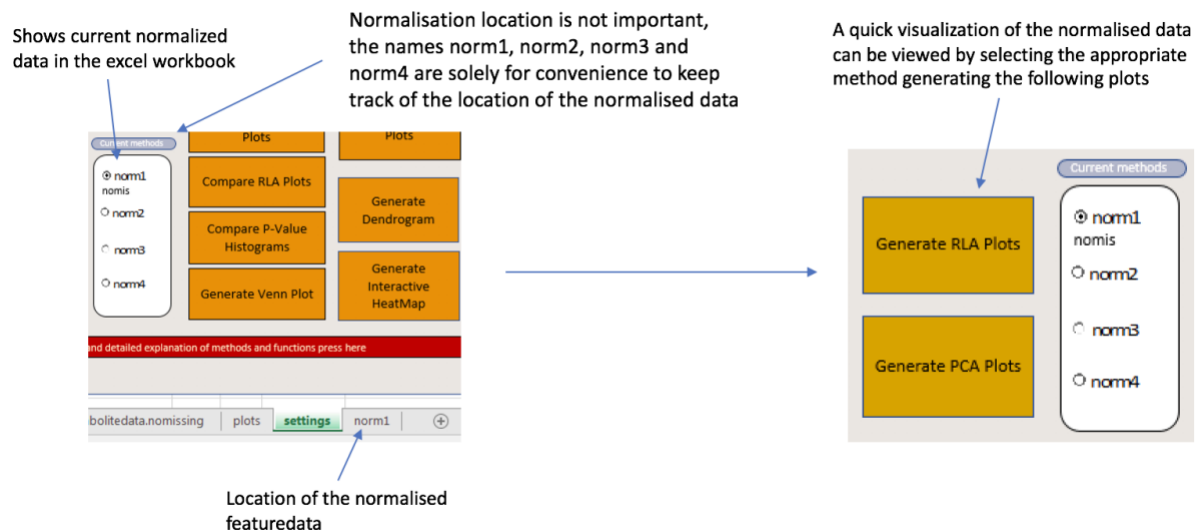




After Clicking the Normalise button the screen asking you where the normalized data is to be saved appears.



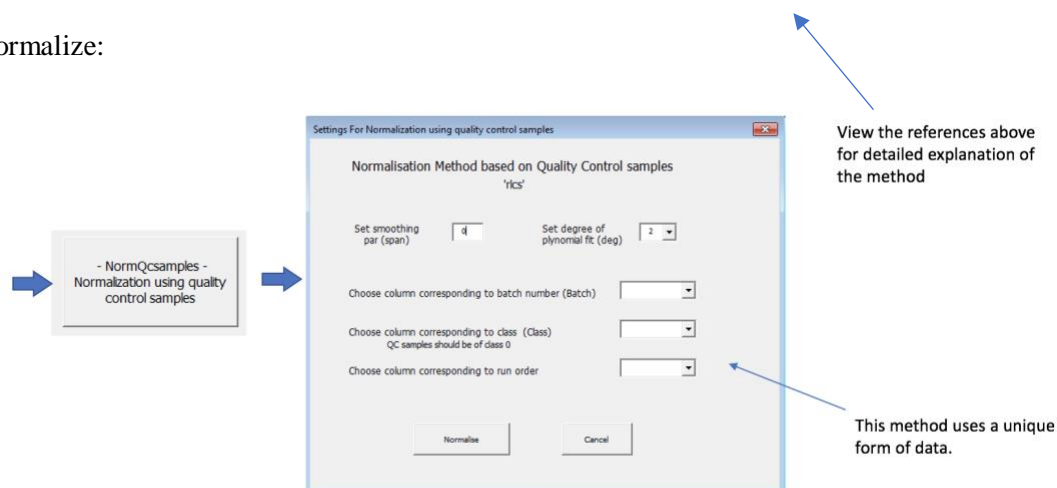
Upon clicking continue, you will return back to the settings sheet but you can notice some changes:



## NormQcsamples

This function is based on the quality control sample based robust LOESS (locally estimated scatterplot smoothing) signal correction (QC-RLSC) method as described by Dunn et al. (2011) and implemented in statTarget (Luan 2017)

To Normalize:

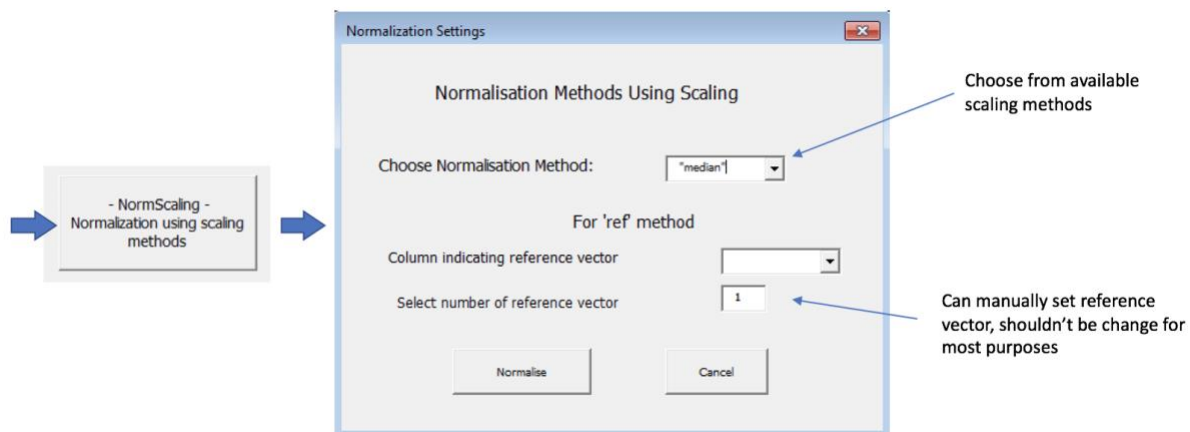


## NormScaling

The scaling normalization methods (Scholz et al. 2004, Wang et al. (2003)) included in the package are normalization to a total sum, normalisation by the median or mean of each sample, and are denoted by *sum*, *median*, and *mean* respectively. The method *ref* normalises the metabolite abundances to a specific reference vector such as the sample weight or volume.

To Normalize:

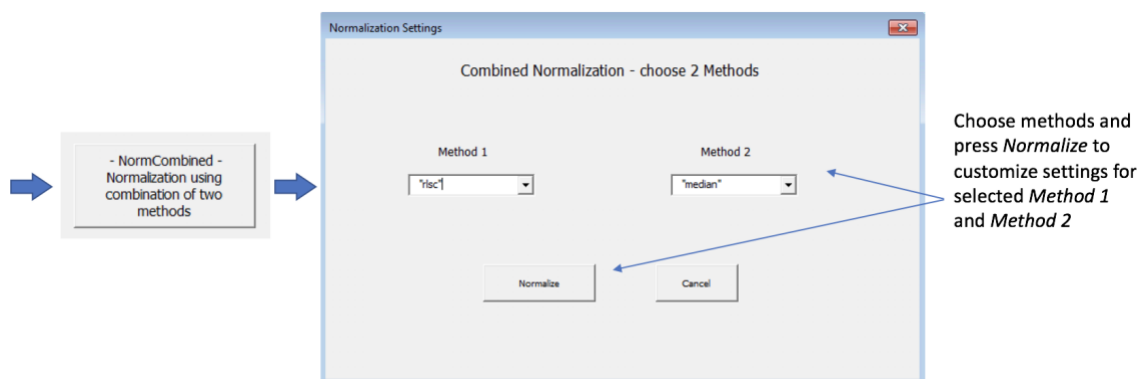




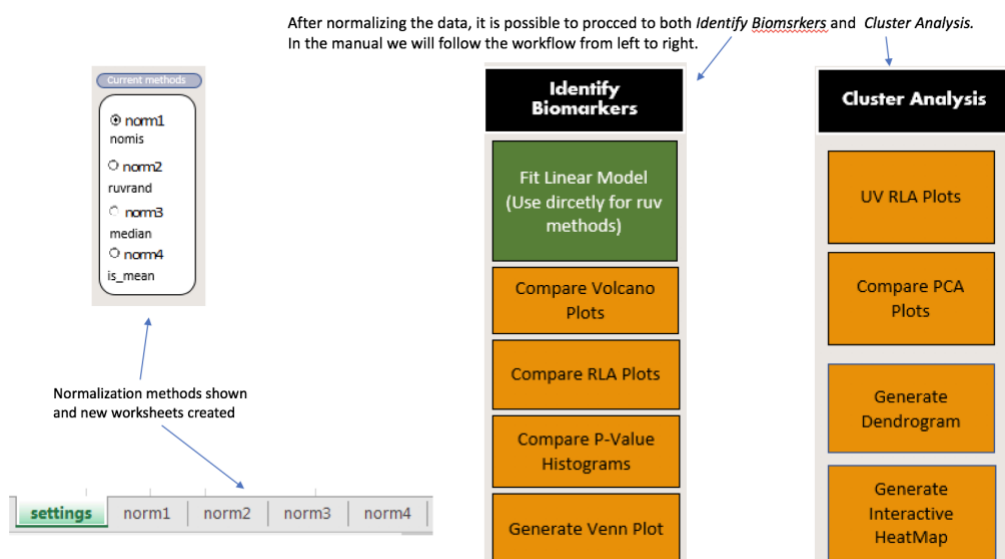
## NormCombined

In some circumstances, researchers use a combination of the above normalizations (i.e., one method followed by another). This can be achieved using the *NormCombined* function. The function defaults to employing 'rlsc' approach followed by the 'median'.

To Normalize:



Note that normalizing the data is not necessary to proceed to fitting a linear model although it is highly recommended to try a few normalization methods when analysis data.

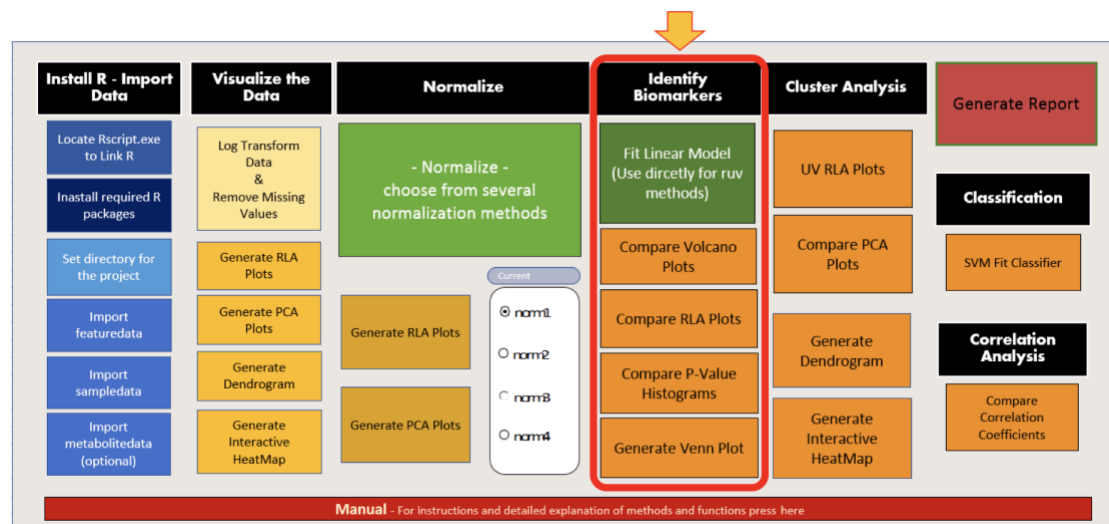


## Assessing and choosing normalization methods

The criteria for assessing and choosing a normalization method implemented NormlizeMets have been described in detail by De Livera et al. (2012), De Livera et al. (2015) and J. A. Gagnon-Bartsch, Jacob, and Speed (2014).

## Identifying Biomarkers

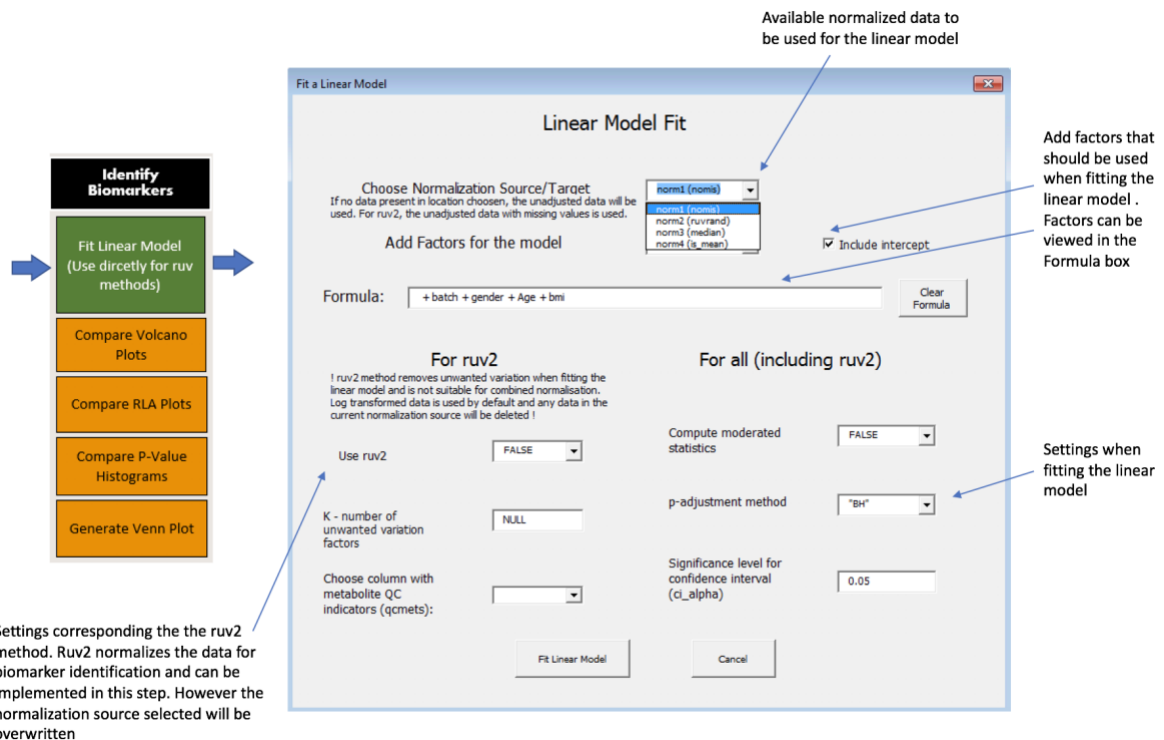
To view and compare the biomarkers identified, first a linear model has to be fitted to the data.



### **Fit Linear Model**

A linear model has to be fitted for every Normalization method that is to be used down the line for Biomarker identification. Setting from one 'run' of the Fit Linear model will be saved for the next.

To Fit Linear Model:



File	Batch	Gender	Age	BMI	Intercept	Batch	Gender	Age	BMI	Intercept	Batch	Gender	Age	BMI	Intercept
1	1	18.0	1.85	0.000000	0.000000	1	1	18.0	1.85	0.000000	1	1	18.0	1.85	0.000000
2	2	19.0	1.90	0.000000	0.000000	2	2	19.0	1.90	0.000000	2	2	19.0	1.90	0.000000
3	3	20.0	2.00	0.000000	0.000000	3	3	20.0	2.00	0.000000	3	3	20.0	2.00	0.000000
4	4	21.0	2.10	0.000000	0.000000	4	4	21.0	2.10	0.000000	4	4	21.0	2.10	0.000000
5	5	22.0	2.20	0.000000	0.000000	5	5	22.0	2.20	0.000000	5	5	22.0	2.20	0.000000
6	6	23.0	2.30	0.000000	0.000000	6	6	23.0	2.30	0.000000	6	6	23.0	2.30	0.000000
7	7	24.0	2.40	0.000000	0.000000	7	7	24.0	2.40	0.000000	7	7	24.0	2.40	0.000000
8	8	25.0	2.50	0.000000	0.000000	8	8	25.0	2.50	0.000000	8	8	25.0	2.50	0.000000
9	9	26.0	2.60	0.000000	0.000000	9	9	26.0	2.60	0.000000	9	9	26.0	2.60	0.000000
10	10	27.0	2.70	0.000000	0.000000	10	10	27.0	2.70	0.000000	10	10	27.0	2.70	0.000000
11	11	28.0	2.80	0.000000	0.000000	11	11	28.0	2.80	0.000000	11	11	28.0	2.80	0.000000
12	12	29.0	2.90	0.000000	0.000000	12	12	29.0	2.90	0.000000	12	12	29.0	2.90	0.000000
13	13	30.0	3.00	0.000000	0.000000	13	13	30.0	3.00	0.000000	13	13	30.0	3.00	0.000000
14	14	31.0	3.10	0.000000	0.000000	14	14	31.0	3.10	0.000000	14	14	31.0	3.10	0.000000
15	15	32.0	3.20	0.000000	0.000000	15	15	32.0	3.20	0.000000	15	15	32.0	3.20	0.000000
16	16	33.0	3.30	0.000000	0.000000	16	16	33.0	3.30	0.000000	16	16	33.0	3.30	0.000000
17	17	34.0	3.40	0.000000	0.000000	17	17	34.0	3.40	0.000000	17	17	34.0	3.40	0.000000
18	18	35.0	3.50	0.000000	0.000000	18	18	35.0	3.50	0.000000	18	18	35.0	3.50	0.000000
19	19	36.0	3.60	0.000000	0.000000	19	19	36.0	3.60	0.000000	19	19	36.0	3.60	0.000000
20	20	37.0	3.70	0.000000	0.000000	20	20	37.0	3.70	0.000000	20	20	37.0	3.70	0.000000
21	21	38.0	3.80	0.000000	0.000000	21	21	38.0	3.80	0.000000	21	21	38.0	3.80	0.000000
22	22	39.0	3.90	0.000000	0.000000	22	22	39.0	3.90	0.000000	22	22	39.0	3.90	0.000000
23	23	40.0	4.00	0.000000	0.000000	23	23	40.0	4.00	0.000000	23	23	40.0	4.00	0.000000
24	24	41.0	4.10	0.000000	0.000000	24	24	41.0	4.10	0.000000	24	24	41.0	4.10	0.000000
25	25	42.0	4.20	0.000000	0.000000	25	25	42.0	4.20	0.000000	25	25	42.0	4.20	0.000000
26	26	43.0	4.30	0.000000	0.000000	26	26	43.0	4.30	0.000000	26	26	43.0	4.30	0.000000
27	27	44.0	4.40	0.000000	0.000000	27	27	44.0	4.40	0.000000	27	27	44.0	4.40	0.000000
28	28	45.0	4.50	0.000000	0.000000	28	28	45.0	4.50	0.000000	28	28	45.0	4.50	0.000000
29	29	46.0	4.60	0.000000	0.000000	29	29	46.0	4.60	0.000000	29	29	46.0	4.60	0.000000
30	30	47.0	4.70	0.000000	0.000000	30	30	47.0	4.70	0.000000	30	30	47.0	4.70	0.000000
31	31	48.0	4.80	0.000000	0.000000	31	31	48.0	4.80	0.000000	31	31	48.0	4.80	0.000000
32	32	49.0	4.90	0.000000	0.000000	32	32	49.0	4.90	0.000000	32	32	49.0	4.90	0.000000
33	33	50.0	5.00	0.000000	0.000000	33	33	50.0	5.00	0.000000	33	33	50.0	5.00	0.000000
34	34	51.0	5.10	0.000000	0.000000	34	34	51.0	5.10	0.000000	34	34	51.0	5.10	0.000000
35	35	52.0	5.20	0.000000	0.000000	35	35	52.0	5.20	0.000000	35	35	52.0	5.20	0.000000
36	36	53.0	5.30	0.000000	0.000000	36	36	53.0	5.30	0.000000	36	36	53.0	5.30	0.000000
37	37	54.0	5.40	0.000000	0.000000	37	37	54.0	5.40	0.000000	37	37	54.0	5.40	0.000000
38	38	55.0	5.50	0.000000	0.000000	38	38	55.0	5.50	0.000000	38	38	55.0	5.50	0.000000
39	39	56.0	5.60	0.000000	0.000000	39	39	56.0	5.60	0.000000	39	39	56.0	5.60	0.000000
40	40	57.0	5.70	0.000000	0.000000	40	40	57.0	5.70	0.000000	40	40	57.0	5.70	0.000000
41	41	58.0	5.80	0.000000	0.000000	41	41	58.0	5.80	0.000000	41	41	58.0	5.80	0.000000
42	42	59.0	5.90	0.000000	0.000000	42	42	59.0	5.90	0.000000	42	42	59.0	5.90	0.000000
43	43	60.0	6.00	0.000000	0.000000	43	43	60.0	6.00	0.000000	43	43	60.0	6.00	0.000000
44	44	61.0	6.10	0.000000	0.000000	44	44	61.0	6.10	0.000000	44	44	61.0	6.10	0.000000
45	45	62.0	6.20	0.000000	0.000000	45	45	62.0	6.20	0.000000	45	45	62.0	6.20	0.000000
46	46	63.0	6.30	0.000000	0.000000	46	46	63.0	6.30	0.000000	46	46	63.0	6.30	0.000000
47	47	64.0	6.40	0.000000	0.000000	47	47	64.0	6.40	0.000000	47	47	64.0	6.40	0.000000
48	48	65.0	6.50	0.000000	0.000000	48	48	65.0	6.50	0.000000	48	48	65.0	6.50	0.000000
49	49	66.0	6.60	0.000000	0.000000	49	49	66.0	6.60	0.000000	49	49	66.0	6.60	0.000000
50	50	67.0	6.70	0.000000	0.000000	50	50	67.0	6.70	0.000000	50	50	67.0	6.70	0.000000
51	51	68.0	6.80	0.000000	0.000000	51	51	68.0	6.80	0.000000	51	51	68.0	6.80	0.000000
52	52	69.0	6.90	0.000000	0.000000	52	52	69.0	6.90	0.000000	52	52	69.0	6.90	0.000000
53	53	70.0	7.00	0.000000	0.000000	53	53	70.0	7.00	0.000000	53	53	70.0	7.00	0.000000
54	54	71.0	7.10	0.000000	0.000000	54	54	71.0	7.10	0.000000	54	54	71.0	7.10	0.000000
55	55	72.0	7.20	0.000000	0.000000	55	55	72.0	7.20	0.000000	55	55	72.0	7.20	0.000000
56	56	73.0	7.30	0.000000	0.000000	56	56	73.0	7.30	0.000000	56	56	73.0	7.30	0.000000
57	57	74.0	7.40	0.000000	0.000000	57	57	74.0	7.40	0.000000	57	57	74.0	7.40	0.000000
58	58	75.0	7.50	0.000000	0.000000	58	58	75.0	7.50	0.000000	58	58	75.0	7.50	0.000000
59	59	76.0	7.60	0.000000	0.000000	59	59	76.0	7.60	0.000000	59	59	76.0	7.60	0.000000
60	60	77.0	7.70	0.000000	0.000000	60	60	77.0	7.70	0.000000	60	60	77.0	7.70	0.000000
61	61	78.0	7.80	0.000000	0.000000	61	61	78.0	7.80	0.000000	61	61	78.0	7.80	0.000000
62	62	79.0	7.90	0.000000	0.000000	62	62	79.0	7.90	0.000000	62	62	79.0	7.90	0.000000
63	63	80.0	8.00	0.000000	0.000000	63	63	80.0	8.00	0.000000	63	63	80.0	8.00	0.000000
64	64	81.0	8.10	0.00											

File

Batch

Gender

Age

BMI

Intercept

Batch

Gender

Age

BMI

Intercept

Batch

Gender

Age

BMI

Intercept

1

2

3

4

5

6

7

8

9

10

11

12

13

14

15

16

17

18

19

20

21

22

23

24

25

26

27

28

29

30

31

32

33

34

35

36

37

38

39

40

41

42

43

44

45

46

47

48

49

50

51

52

53

54

55

56

57

58

59

60

61

62

63

64

1

2

3

4

5

6

7

8

9

10

11

12

13

14

15

16

17

18

19

20

21

22

23

24

25

26

27

28

29

30

31

32

33

34

35

36

37

38

39

40

41

42

43

44

45

46

47

48

49

50

51

52

53

54

55

56

57

58

59

60

61

62

63

64

1

2

3

4

5

6

7

8

9

10

11

12

13

14

15

16

17

18

19

20

21

22

23

24

25

26

27

28

29

30

31

32

33

34

35

36

37

38

39

40

41

42

43

44

45

46

47

48

49

50

51

52

53

54

55

56

57

58

59

60

61

62

63

64

1

2

3

4

5

6

7

8

9

10

11

12

13

14

15

16

17

18

19

20

21

22

23

24

25

26

27

28

29

30

31

32

33

34

35

36

37

38

39

40

41

42

43

44

45

46

47

48

49

50

51

52

53

54

55

56

57

58

59

60

61

62

63

64

1

2

3

4

5

6

7

8

9

10

11

12

13

14

15

16

17

18

19

20

21

22

23

24

25

26

27

28

29

30

31

32

33

34

35

36

37

38

39

40

41

42

43

44

45

46

47

48

49

50

51

52

53

54

55

56

57

58

59

60

61

62

63

64

1

2

3

4

5

6

7

8

9

10

11

12

13

14

15

16

17

18

19

20

21

22

23

24

25

26

27

28

29

30

31

32

33

34

35

36

37

38

39

40

41

42

43

44

45

46

47

48

49

50

51

52

53

54

55

56

57

<

## Compare Volcano Plots:

Metabolites above this will be colored differently

Identify Biomarkers

Fit Linear Model (Use directly for ruv methods)

Compare Volcano Plots

Compare RLA Plots

Compare P-Value Histograms

Generate Venn Plot

Volcano Plots Settings

Compare Volcano Plots

Choose up to 4 methods for comparison - For a regular volcano Plot, only choose one method  
Only available data that has a linear model fitted to it (using the Linear Model Fit function) will appear in the options

Plot 1: norm1 (norm1) Plot 2: norm2 (ruvrand) Plot 3: norm3 (median) Plot 4: norm4 (ruv2)

P - limit: 0.05 y - range (optional): min max

coefficient: 1 Negative control column: neg\_controls

x - label: "Coefficients" Positive control column: pos\_controls\_1

y - label: "-log10(p-value)" Plot Name: "InteractiveVolcanoP"

Factor to use for the plots: bmi

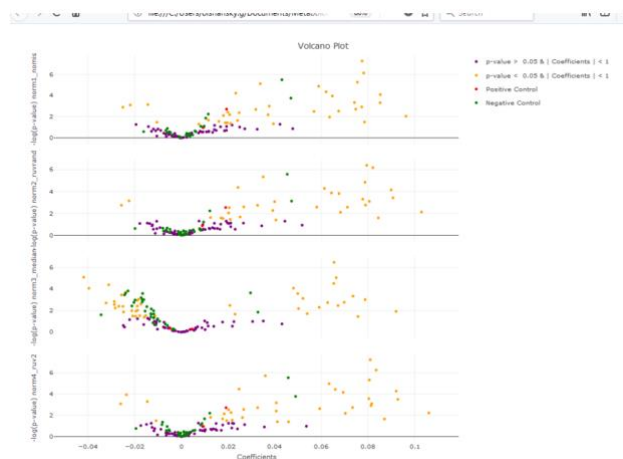
Generate Plots Cancel

Optionally specify the y-range of the plots (the same range will be used for all plots)

Controls will be colored differently

Coefficients for the plots will be for that factor

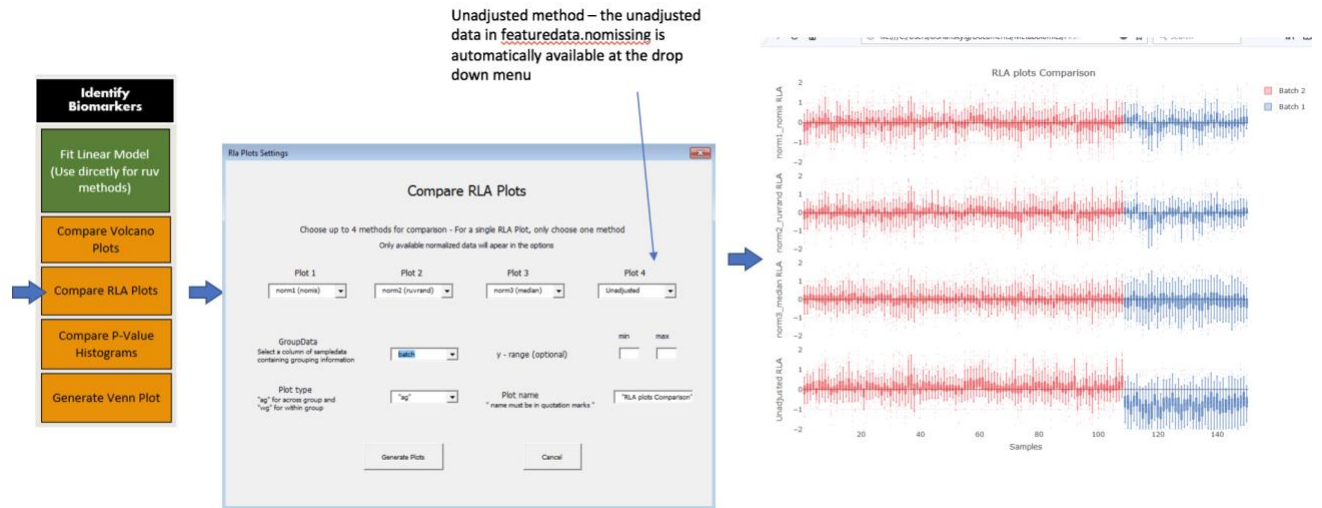
The plot is saved in the working directory and opens in the default browser.



## Compare RLA plots

Used to assess normalization by comparing relative log abundance plots, similar input to the *Generate RLA plots* function

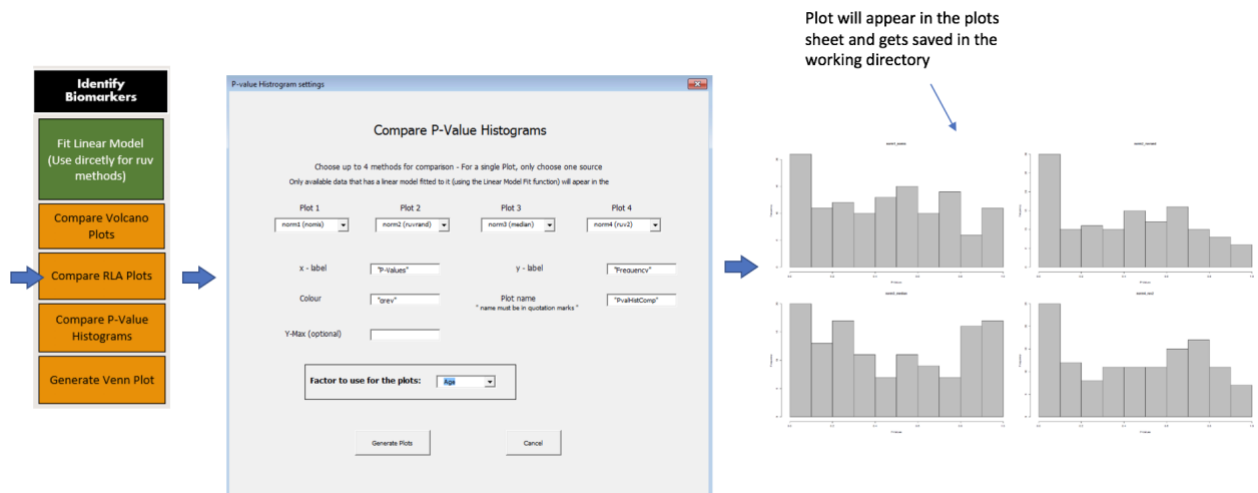
Compare Rla Plots:



## Compare P-Value Histograms

Compare histograms of the coefficient's p-values. The distribution of the p-values should be used to assess the success of the normalization.

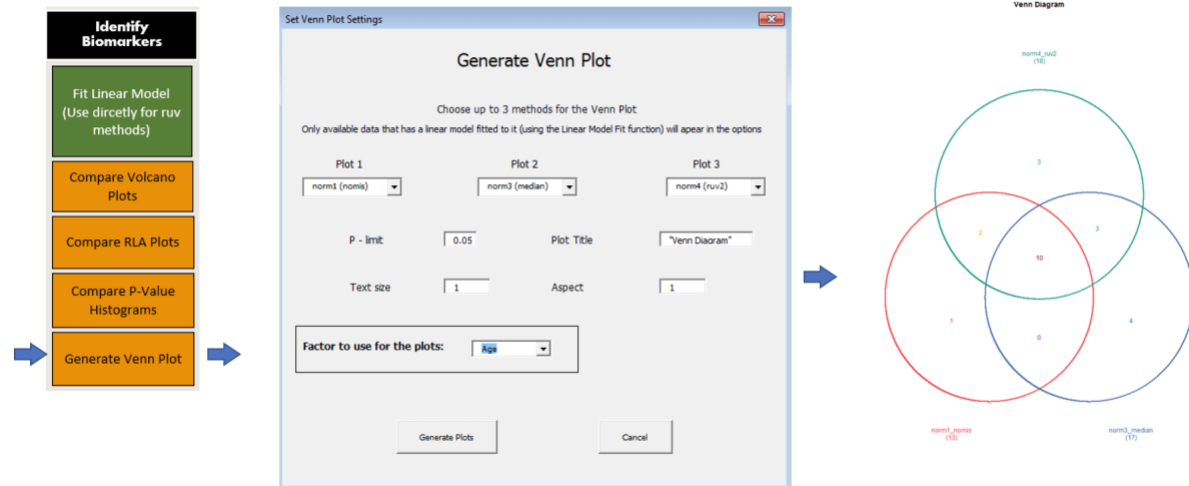
Compare P-Value histograms:



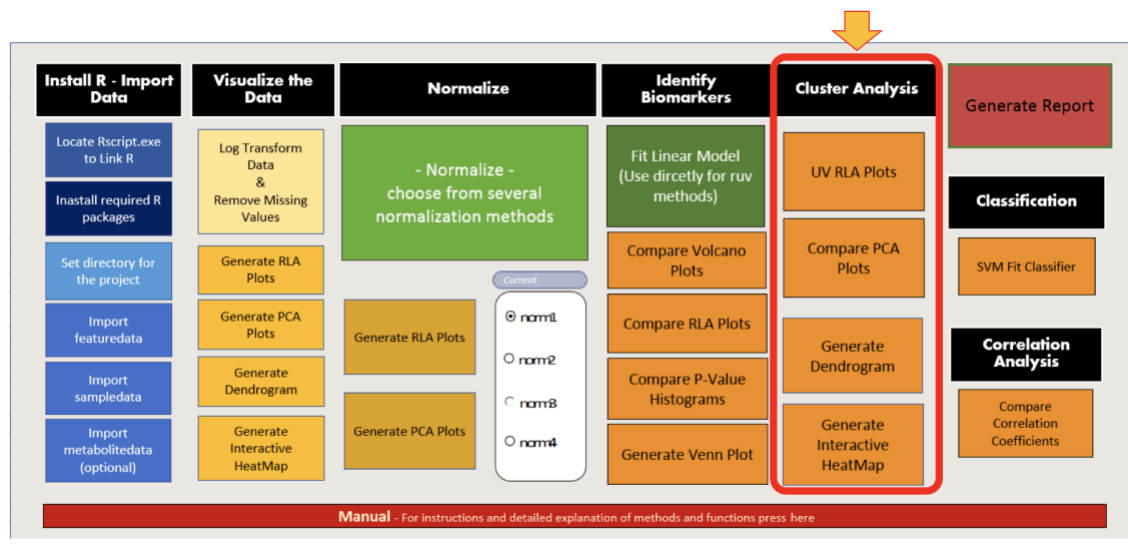
## Generate Venn Plot

Generates a Venn plot that compares the biomarkers identified by the different normalization methods.

Generate Venn Plot:

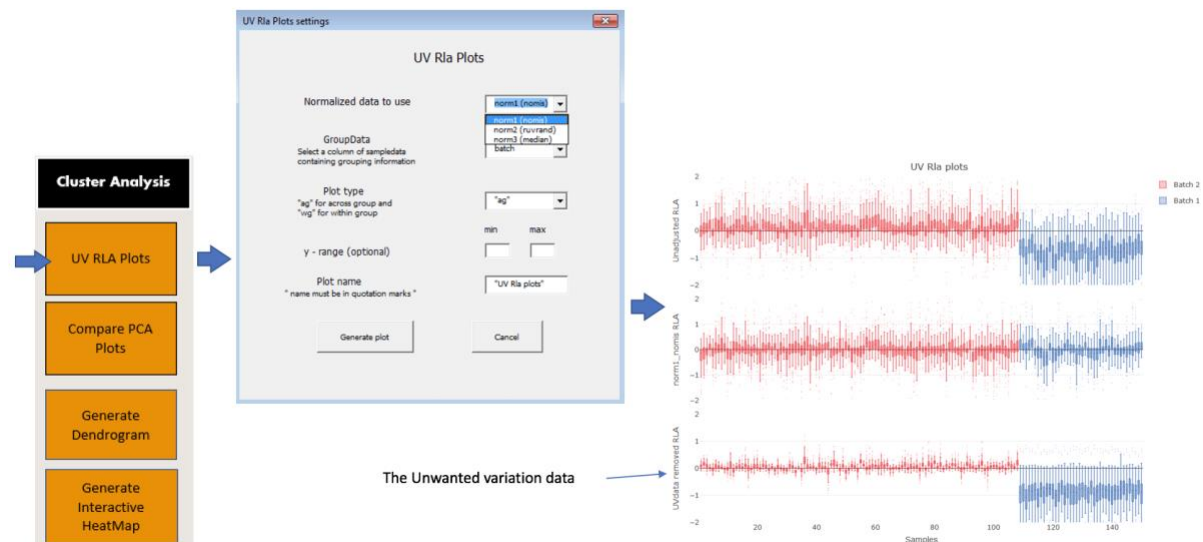


## Cluster Analysis



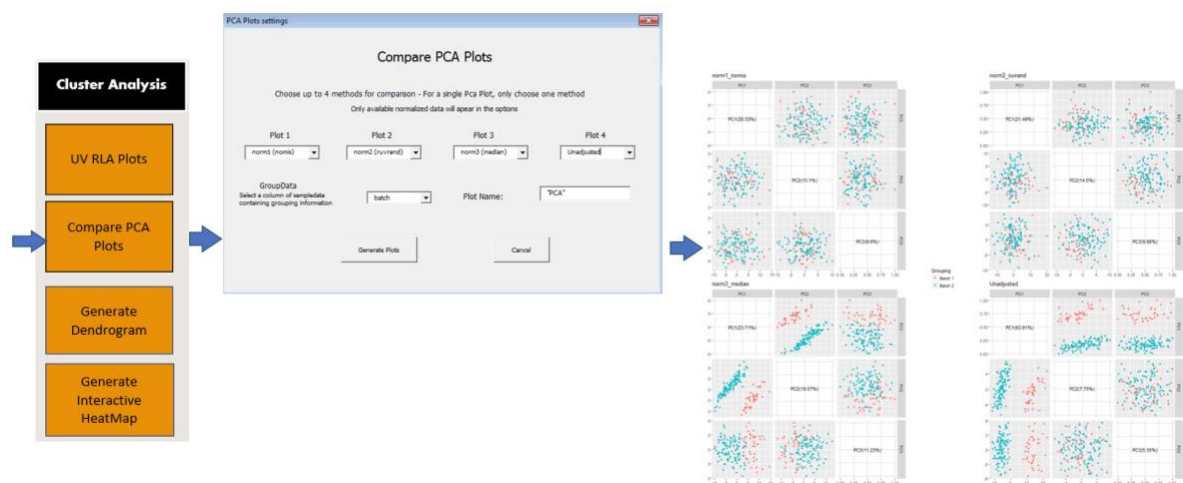
## UV RLA plots

Unwanted Variation relative log abundance plots enable visualisation of the unwanted variation removed by each normalization method.



## Compare PCA Plots

Compare principal component multi-plots for differed normalization methods.



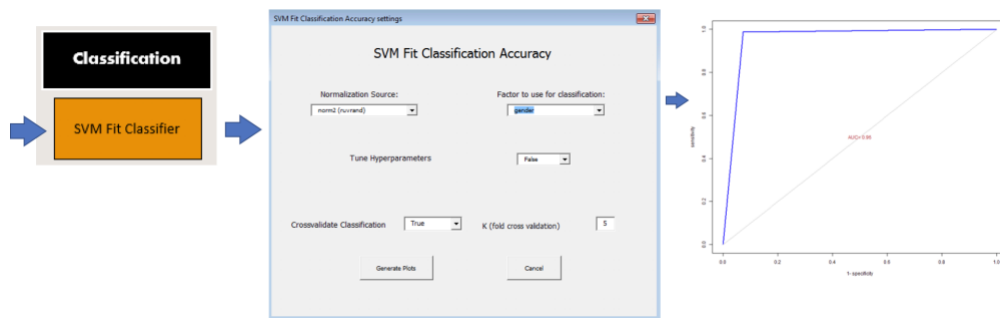
**Generate Dendrogram and Generate Interactive HeatMap** are identical to those discussed in the *Visualize the Data* section. The user has to choose the normalized data to be used.

## Classification

Classification accuracy is a good way to assess the success of a given normalization method

### SVM-Fit

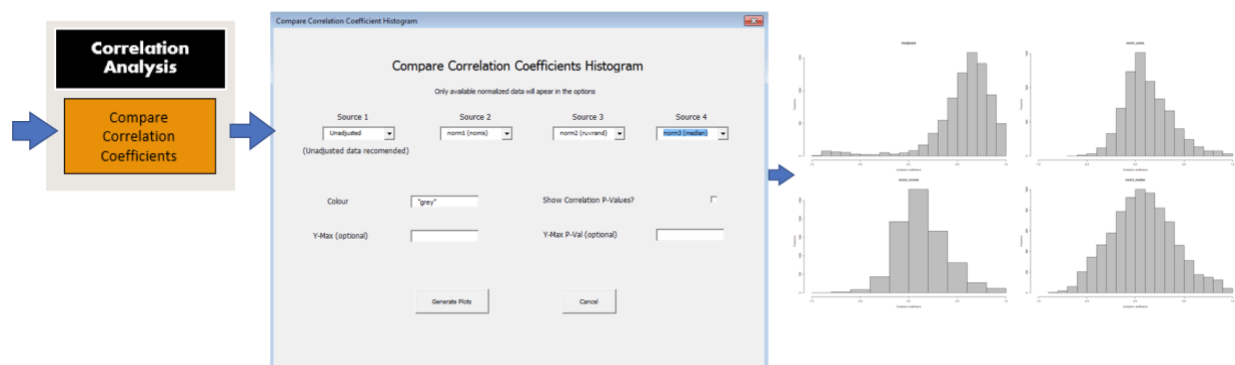
The Support Vector Machines method is used to classify the data, the classification accuracy is then assessed based on the factor specified.



## Correlation Analysis

It is important to look at correlation coefficients when normalizing data.

### Compare Correlation coefficients

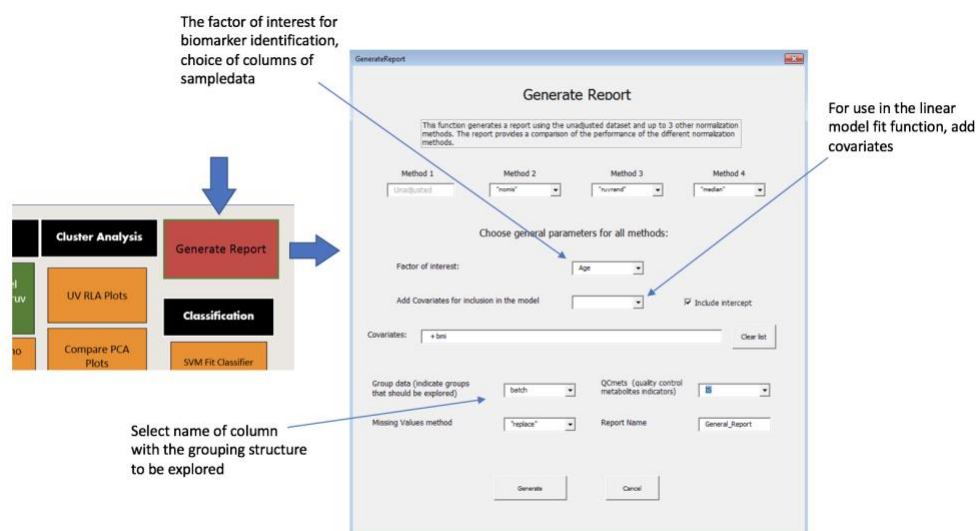


## Generate Report

The *Generate Report* function generates an interactive report based on basic user input. There is a choice of up to 3 normalization methods to be included together with the unadjusted data. The report includes various plots and diagnostic to assess the normalization. Guidance on interpretation of the various plots, together with notes of what the user should look for when assessing the results is provided in the generated document.

Generate Report:





An example report is available in your downloaded ExNormalizeMets docs folder.

## **Full Package Vignette:**

For package vignette with detailed explanations of methods and workflow, follow the link to:  
[https://cran.r-project.org/web/packages/NormalizeMets/vignettes/NormalizeMets\\_vignette.html](https://cran.r-project.org/web/packages/NormalizeMets/vignettes/NormalizeMets_vignette.html)