How to use LIMMA statistical analysis code?

The main code is present in Fig_1B**limma_main_Fig1B.R**. Run the code in RStudio version >= 1.3.959 (and R version >= 4.0.2), The code will ask for a series of input as listed below. To reproduce Fig1B, provide the exact inputs depicted in blue color:

1. Enter treatment name(case insensitive) as it appeared in the iBAQ/LFQ column= AROM
2. Enter control name(case insensitive) as it appeared in the iBAQ/LFQ column= WT
3. Enter 1 for iBAQ or 0 for LFQ= 0
4. Enter the number of treatment replicates=3
5. Enter the number of control replicates=3
6. Enter the fold change cut off=2
7. Enter a filename for final data= Fig1B_final_data
8. Enter a filename for limma plot= Fig1B_limma_plot

After a successful run, **Results_Fig1B** folder will be created inside Fig_1B folder. The results folder will contain following output files:

1. Fig1B_final_data.tsv will contain data for Fig1B
2. Fig1B_limma_plot.html will be equivalent to Fig1B (see Important Note)
3. Exclusive_proteins_AROM.tsv will contain exclusive proteins in AROM+
4. Exclusive_proteins_WT.tsv will contain exclusive proteins in WT

Important Note:

1. Since the missing values are imputed randomly (from normal distribution), users can see minute change in numbers associated with fold change and p-values across multiple runs of the pipeline. However, the major patterns in the data do not change.
2. Run the code by clicking Source button in RStudio, running line by line individually is NOT recommended.

How to use the MLP scoring code?

1. The main code is present in AROM_Paper_Codes\MLP_Scoring**MLP_Main_Code.m**.
2. Helper functions are present in AROM_Paper_Codes\MLP_Scoring**MLP_Helper_Functions.m**.
3. We will use peptide (identified using IMS) masses in **Sample_Data_For_MLP** folder to demonstrate the algorithm and henceforth we will refer those masses as IMS masses.
4. Running ***MLP_Main_Code.m*** will create two folders: *AROM_Paper_Codes\MLP_Scoring\Sample_Data_For_MLP**Output*** and AROM_Paper_Codes\MLP_Scoring\Sample_Data_For_MLP**mass_lists_after_Deisotoping**.
5. ***mass_lists_after_Deisotoping*** folder will contain the IMS mass list after deisotoping.
6. For each IMS mass in a deisotoped list (IMS mass file), ***MLP_Main_Code.m*** looks for all the matching LC-MS peptides within a mass tolerance and the folder *AROM_Paper_Codes\MLP_Scoring\Sample_Data_For_MLP**Output*** will contain the results for each such IMS mass file.
7. However, as discussed in the paper that at this stage we face an ambiguity, i.e. there is a possibility that one IMS mass can map to multiple LC-MS peptides (even originating from different proteins).
8. In order to avoid this ambiguity we used MLP scoring (see *Analysis and identification of peptide masses* section of the paper) and the results are stored in *AROM_Paper_Codes\MLP_Scoring\Sample_Data_For_MLP\Output**MLP*** folder, information of only those LCMS peptides that has the best and second best MLP scores are retained here.
9. We then use the Unprot ids of these peptides to perform GO and Pathway (using Reactome DB) enrichment analyses as discussed in the next section.

How to use the Gene Ontology (GO) and Reactome pathway analyses code?

We perform Gene Ontology analysis and Reactome pathway analyses as described in the following links

GO analysis
Reactome pathway analysis
The code location and structure are explained in the following steps:

1. R code for all of these analyses use the output after MLP scoring.
2. The folder *AROM_Paper_Codes**Fig_1_CD*** contains code and data for pathway enrichment analysis to reproduce the Figures 1C-D. Running *Pathway_Enrichment_Analysis_Fig_1CD.R* will generate excel files corresponding to Figures 1C-D in the folder *AROM_Paper_Codes\Cytoscape_Data_Generation**Sample_Data_for_pathway_analysis***.
3. Similarly, the folder *AROM_Paper_Codes\Fig_2_ABCD* contains code and data for pathway enrichment analysis to reproduce the Figures 2A-D. Running *Pathway_Enrichment_Analysis_Fig_2ABCD.R* will produce excel files corresponding to Figures 2A-D in the folder *AROM_Paper_Codes\Cytoscape_Data_Generation**Sample_Data_for_pathway_analysis***.
4. The folder *AROM_Paper_Codes**Fig_5*** contains code and data for GO enrichment analysis showed in Figures 5B-D. Running, *Fig_5_GO_Enrichment_Analysis.R* will yield excel files corresponding to Figures 5B-D in the folder *AROM_Paper_Codes\Cytoscape_Data_Generation**Sample_Data_for_GO_analysis***.
5. The R session information for each of the above analyses are provided in the respective folders.

How to use the Cytoscape data generation code?

1. We used Cytoscape in order to improve the graphics produced by the R codes in the previous section. The data structure yielded by functions from ClusterProfiler and ReactomeDB R packages was not compatible with the Cytoscape. Therefore, we wrote 2 MATLAB scripts to produce Cytoscape compatible data structure from the R data structure:
    I. *AROM_Paper_Codes\Cytoscape_Data_Generation**R2cytoscape_GO_data.m*** - This script uses data from the folder *AROM_Paper_Codes\Cytoscape_Data_Generation**Sample_Data_for_GO_analysis*** and creates Cytoscape compatible edge tables in the folder *AROM_Paper_Codes\Cytoscape_Data_Generation**Cytoscape_NW_Tab_GO*** which can then be loaded into Cytoscape and figures equivalent to Figures 5B-D can be created.
    II. *AROM_Paper_Codes\Cytoscape_Data_Generation**R2cytoscape_Pathway_data.m*** - This script uses data from the folder *AROM_Paper_Codes\Cytoscape_Data_Generation**Sample_Data_for_pathway_analysis*** and creates Cytoscape compatible edge tables in the folder *AROM_Paper_Codes\Cytoscape_Data_Generation**Cytoscape_NW_Tab_Pathway*** which can then be loaded into Cytoscape and figures equivalent to Figures 1C-D and 2A-D can be created.

Important Note:

We used MATLAB 2019b for MLP scoring and Cytoscape data generation codes.

Tips:

1. If you encounter difficulty installing package *vctrs* then, download and install it from the source. Example, install.packages('C:\Users\W.Aftab\Downloads\vctrs_0.3.4.tar.gz', repos = NULL, type = "source")
2. You might need to install 'DO.db' separately by using this command: BiocManager::install('DO.db')
3. If you encounter difficulty installing package *tidyr* then, download and install it from the source. Example, install.packages('C:\Users\W.Aftab\Downloads\tidyr_1.1.2.tar.gz', repos = NULL, type = "source")