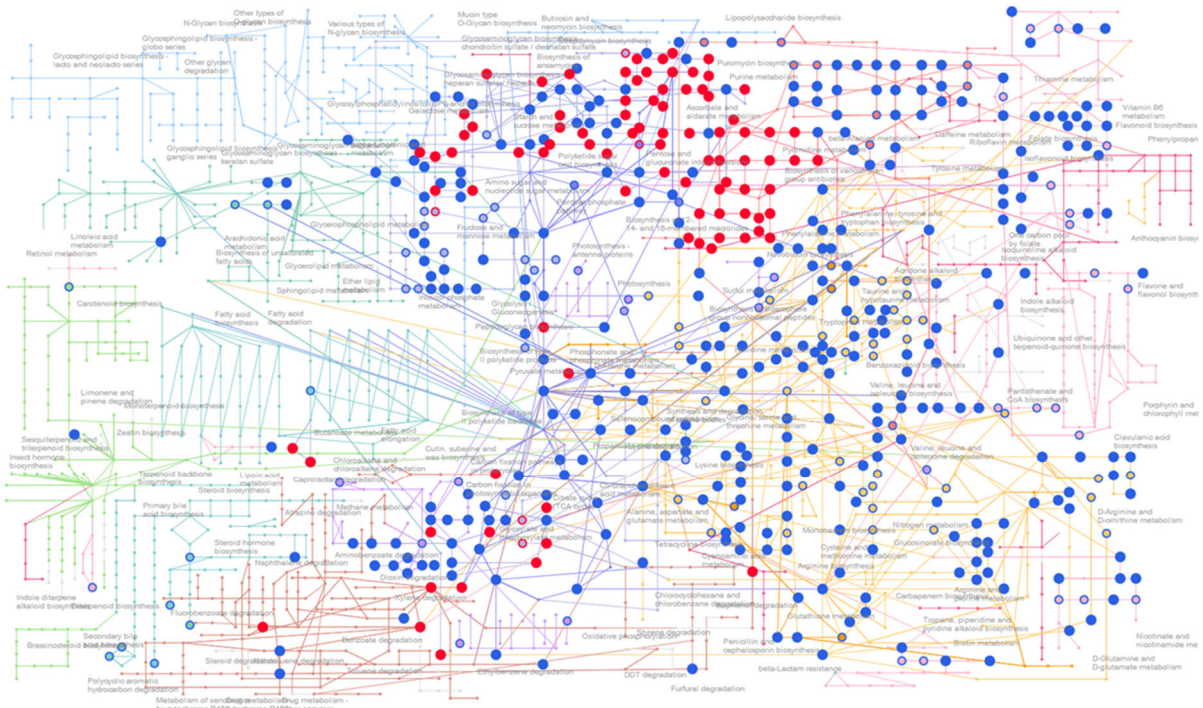
Part 2 (Exercises 2,3 and 4):

Pathways Analysis and Multiomics



3rd Oxford Metabolomics Data Processing and Analysis Workshop

Overview

In **Exercise-1** you performed data processing and statistical analysis of an untargeted metabolomics dataset. In this **Exercise** you will perform pathways analysis (targeted and untargeted) and multiomics data integration using MetaoAnalyst with data from the same metabolomics experiment. The aim is to interpret the results in a metabolic pathway context. This exercise will guide you through the process of putting your statistical analysis into biological context and to produce a report at the end via MetaboAnalyst. You will also be asked a series of questions to help with understanding the results.

As in exercise 1 we will follow a step by step guide which includes screen shots demonstrating how to perform the analysis and display the results. If you are already familiar with using MetaboAnalyst for Pathways Analysis and do not want to be guided by screenshots you can go straight to the final section of this document which provide a list of tasks and questions to be answered. It is expected the majority of people on the course will benefit from the step by step guide.

If you get stuck or have any questions along the way please put up your hand and those running the course will come and help.

To complete Exercise 2, you will need:

1. Access to MetaboAnalyst online: https://www.metaboanalyst.ca/MetaboAnalyst/home.xhtml
2. There are three different data files for Part 2:
    - Targeted Pathways Analysis: '**Exercise_2_DATA_MA.csv' (Part 1)**
    - Untargeted Pathways Analysis: '**Exercise-3_untargeted.csv' (Part 2)**
    - Multiomics (metabolomics and transcriptomics): '**Exercise-4_multiomics (Part 3)**
3. These instructions to follow.

Datafiles are available on the Sharepoint for the workshop. Please make sure you have downloaded a local copy of these data files to your computer before starting the exercise. You may also find it useful to have a hard copy of this exercise sheet or have it open on a separate screen when using MetaboAnalyst.

Part 1: Targeted Pathways Analysis

Step by step guide (with screen shots)

Please follow the step by step guide below to analyse the dataset you are given and create a Pathways analysis report in MetaboAnalyst. You will then have some questions to answer about the results (if you are an experienced MetaboAnalyst user you can go straight the list of talks which are given on the last page of this document).

Part 1: Targeted Pathways Analysis

1. Copy the .csv file called **'Exercise_2_DATA_MA.csv'** to your local computer (this data file will have been sent to you by email and will also have been uploaded to the Teams site for the workshop).

2. Open MetaboAnalyst https://www.metaboanalyst.ca/

3. Select: **>>click here to start<<** which will open up the pyramid of modules (see Figure 1 below).

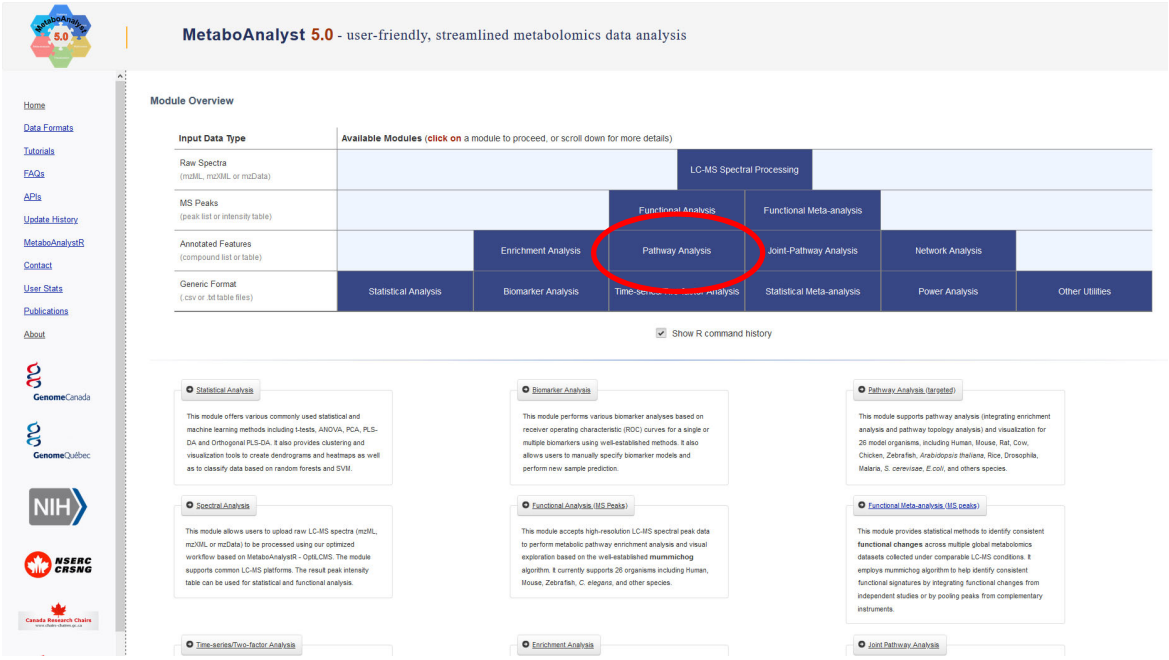


Figure 1: Pyramid of MetaboAnalyst modules.

4. Select **'Pathways Analysis'** which will open up a window as in Figure 2 below. Click on '**A concentration table'** as circled in Figure 2.

5. Complete as shown in Figure 3. Browse for the file called **'Exercise_2_DATA_MA.csv'** you just downloaded.

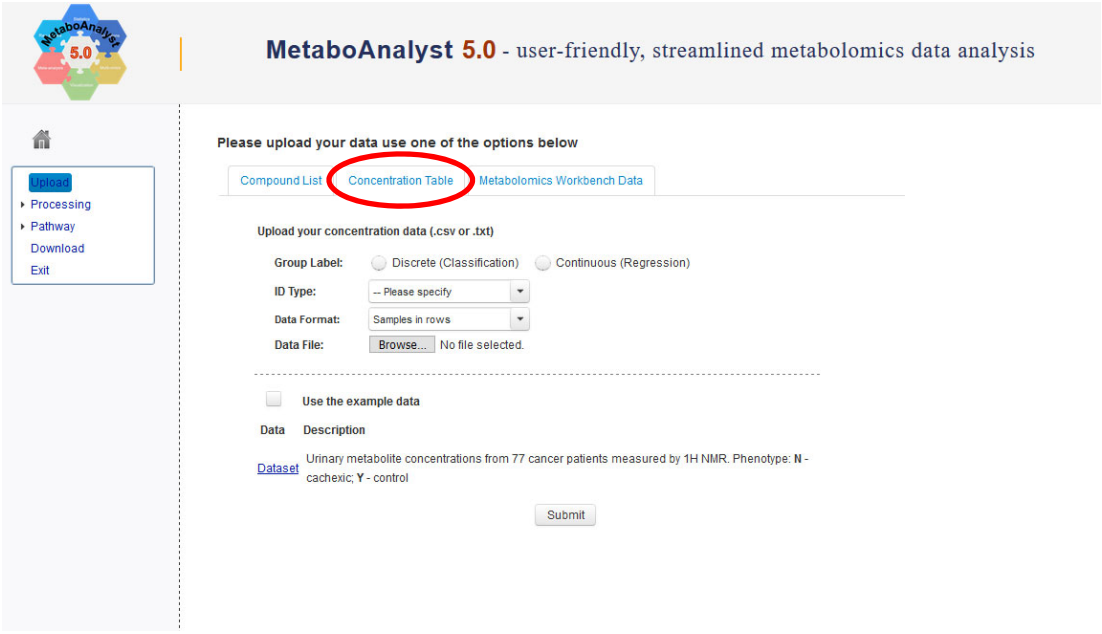


Figure 2: Select 'concentration table'.

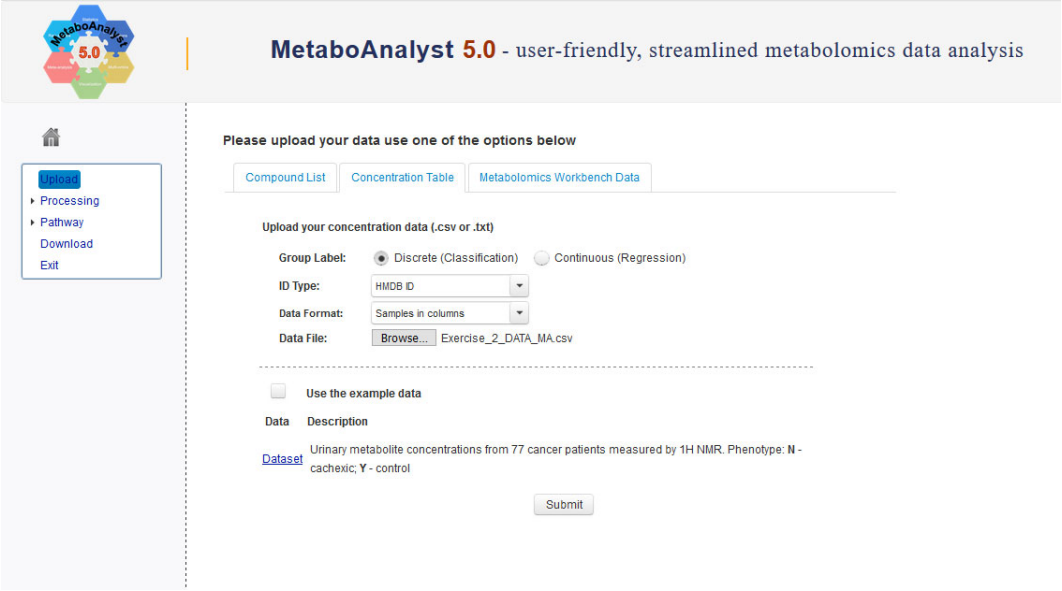


Figure 3: Complete details as shown, browse for the file called 'Exercise_2_PathwaysAnalysis_MA' you just downloaded and click 'submit'

6. You are then taken to a data integrity check page (Figure 4).

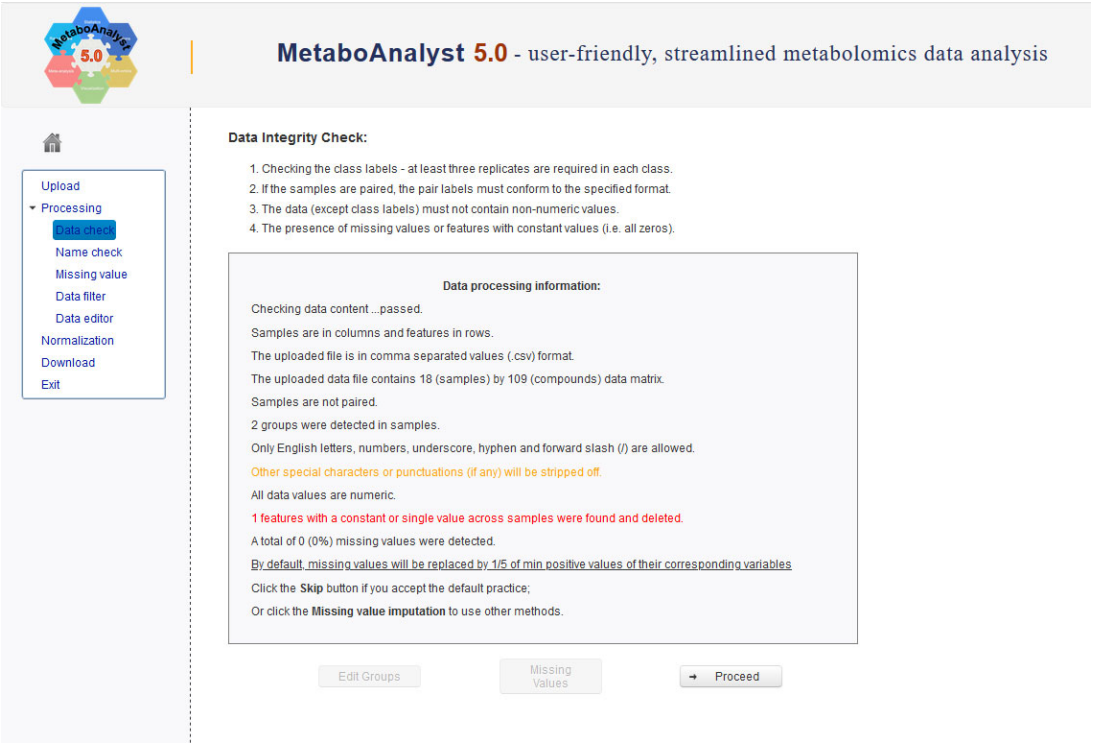


Figure 4: Data Integrity Check page.

7. Click 'Proceed' at the 'Data integrity' page.

8. Check the **'Compound Name/ID Standardization'** page (Figure 5). If any of the naming of metabolites is not correct a 'name check' page will appear. There is a chance to update the name at this point of simply continue and those queried in red will be ignored in the pathways analysis (N.B. 2 are queried in red but we will continue without amending them).

9. Once ready click on 'submit' at the bottom of the list. This opens up the 'Normalisation overview' page familiar from Exercise-1 on Data Processing.

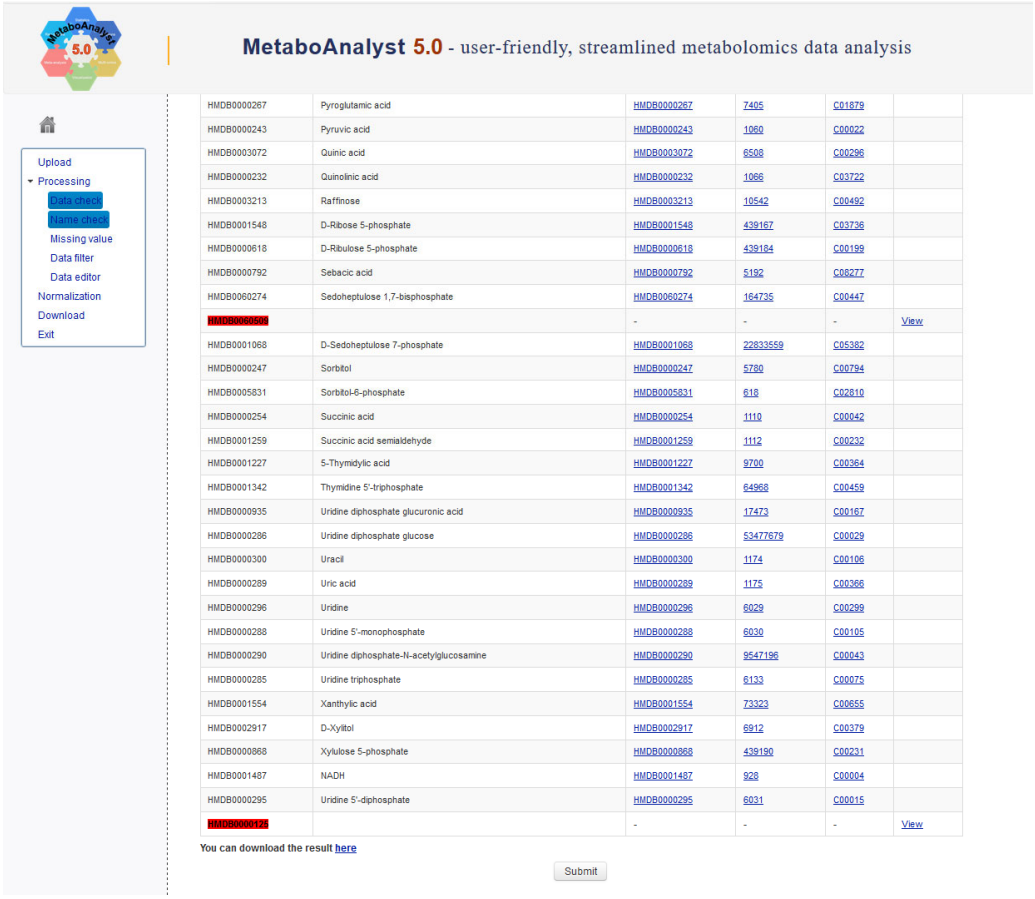


Figure 5: Compound Name/ID Standardization' page.

10. Select 'Normalisation by Sum', 'Log' for data transformation and 'Pareto Scaling' under Data Scaling click on 'Normalise' and then 'Proceed'. (Figure 6)

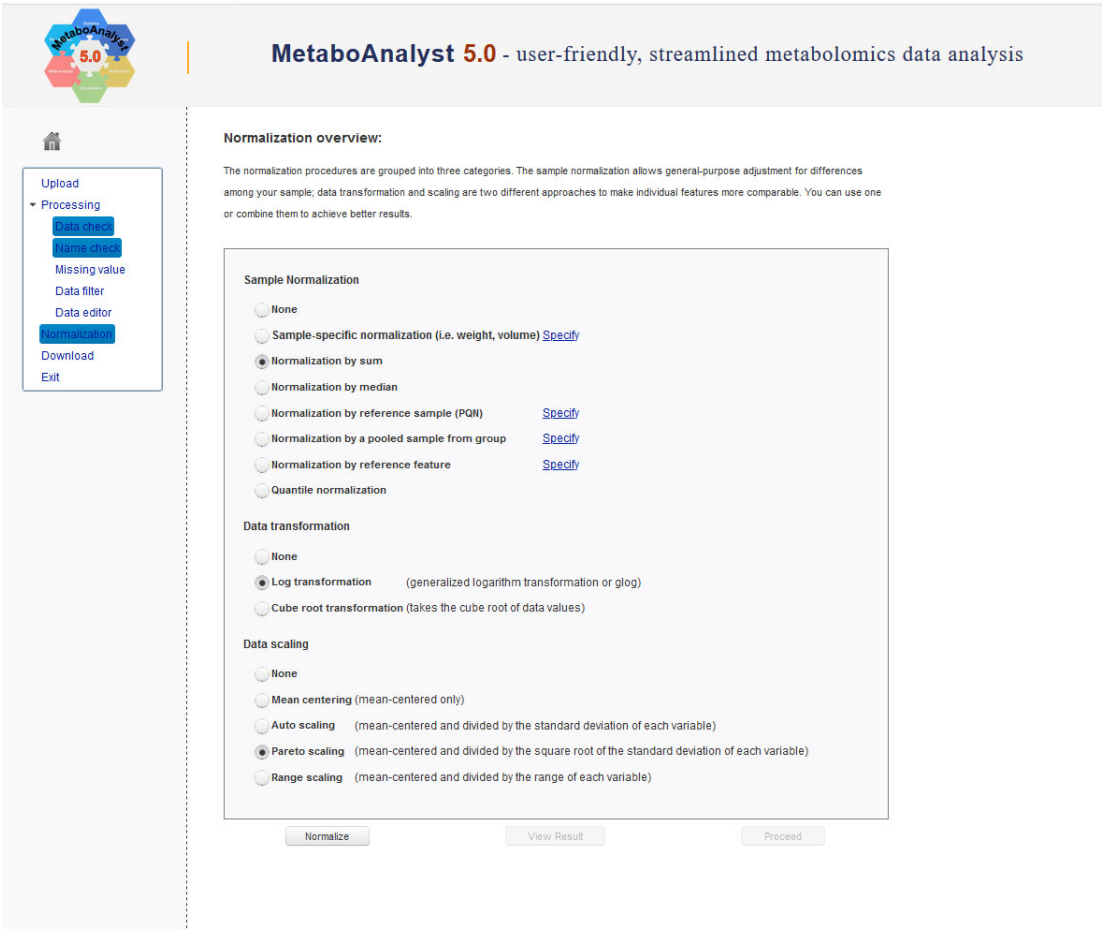


Figure 6: Data normalisation page (select the same normalisation, transformation and scaling approaches optimised from the statistical analysis of the data then click 'Proceed').

11. On the next page leave '**Specify Pathways analysis algorithms'** as default (Scatter plot, Global test, Relative betweenness centrality and Use all compounds…) and select the 'Pathway Library' that is relevant to the samples being analysed (default is Homo sapiens/KEGG which us suitable for human derived cells, tissues and bio-fluids and will be used here).

12. Ensure 'Homo sapiens (KEGG)' is selected (Figure 7). Select 'Submit'

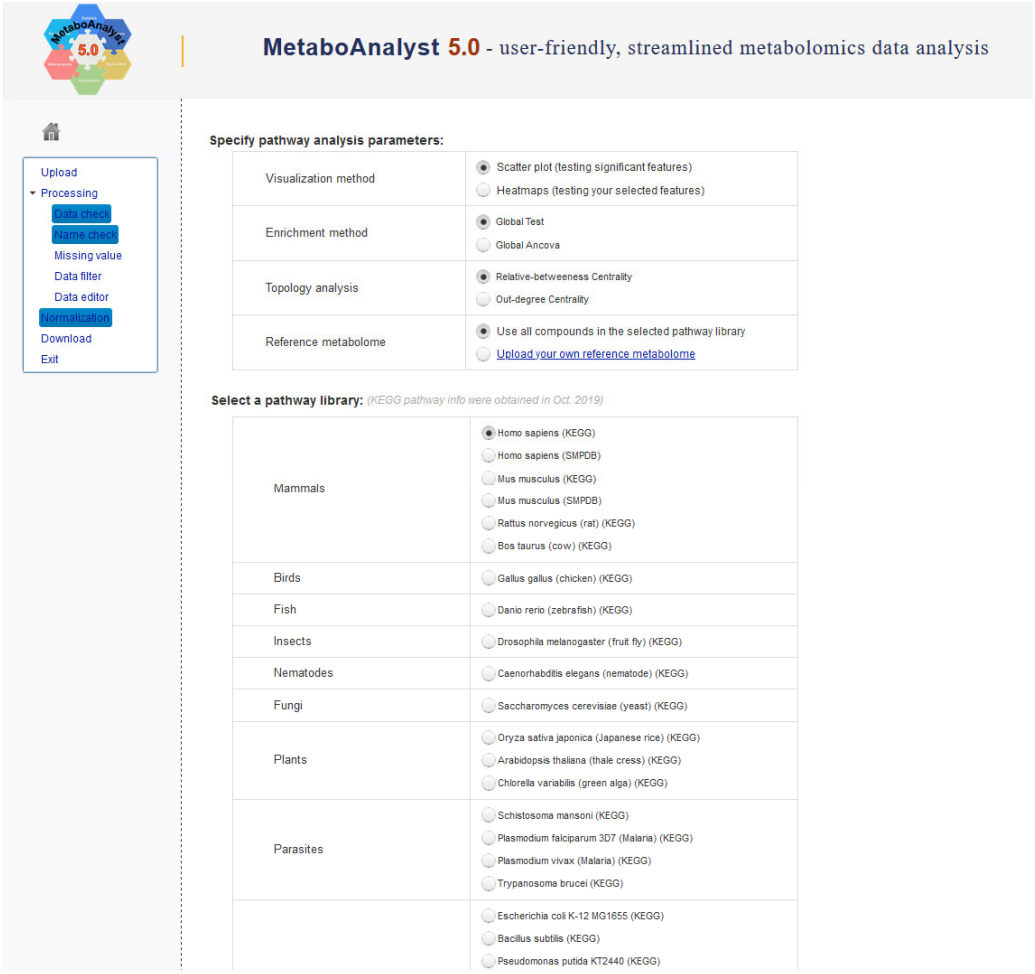


Figure 7: Specify Pathways Analysis parameters.

13. The next view will show the results of the Pathways Analysis (Figure 8). On the left hand side is an overview of the Pathways identified and the right hand side shows metabolites in those pathway. This is displayed when the circles in the graph are clicked on.

    [About compound colours within the pathway - light blue means those metabolites are not in your data and are used as background for enrichment analysis; grey means the metabolite is not in your data and is also excluded from enrichment analysis (only applicable if you have uploaded a custom metabolome profile); other colours (varying from yellow to red) means the metabolites are in the data with different levels of significance.]

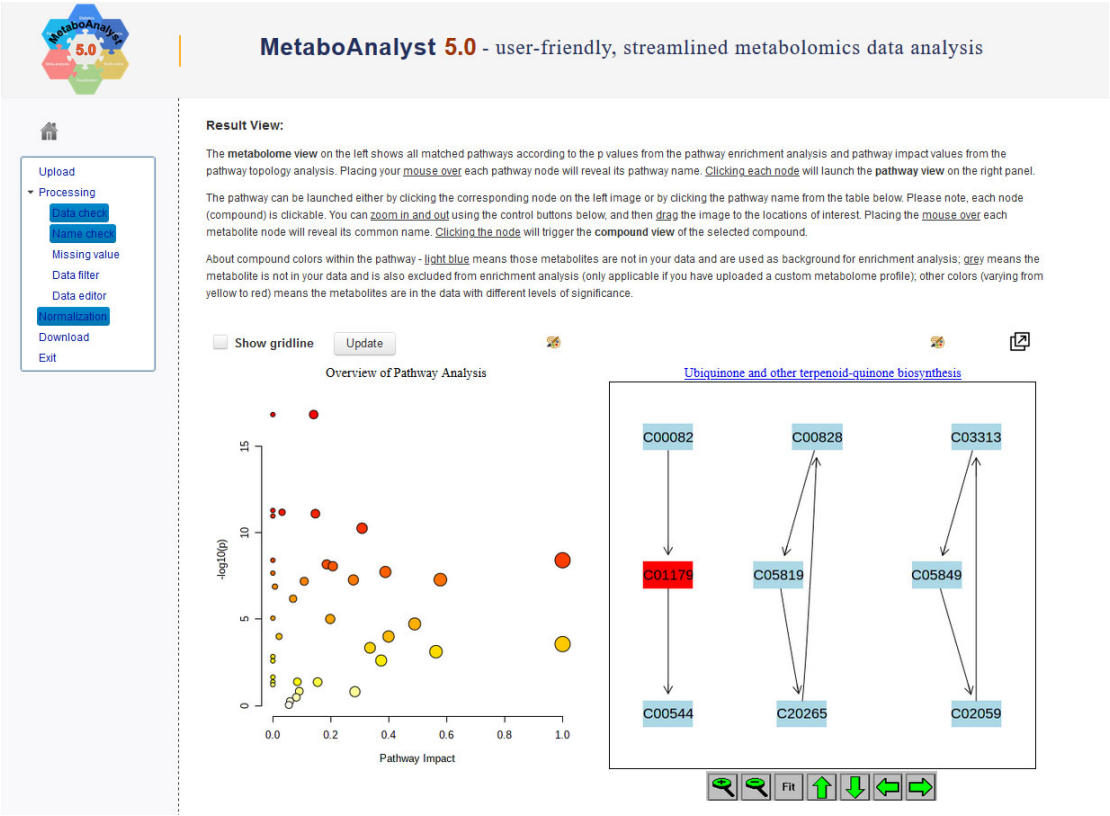
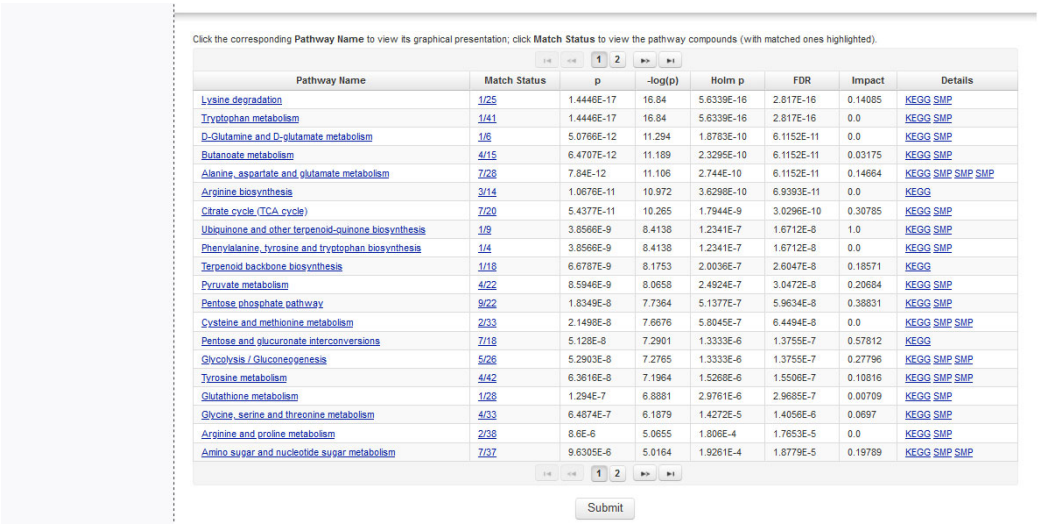


Figure 8: Pathways Analysis results screen

14. Below these graphs is a list of the pathways identified as being modified along with statistics associated with the significance of these modifications (Figure 9).



Pathway Name	Match Status	p	-log(p)	Holm p	FDR	Impact	Details
Lysine degradation	1/25	1.4446E-17	16.84	5.6339E-16	2.817E-16	0.14085	KEGG SMP
Tryptophan metabolism	1/41	1.4446E-17	16.84	5.6339E-16	2.817E-16	0.0	KEGG SMP
D-Glutamine and D-glutamate metabolism	1/6	5.0766E-12	11.294	1.8783E-10	6.1152E-11	0.0	KEGG SMP
Butanoate metabolism	4/15	6.4707E-12	11.189	2.3295E-10	6.1152E-11	0.03175	KEGG SMP
Alanine, aspartate and glutamate metabolism	7/28	7.84E-12	11.106	2.744E-10	6.1152E-11	0.14664	KEGG SMP SMP SMP
Arginine biosynthesis	3/14	1.0676E-11	10.972	3.6298E-10	6.9393E-11	0.0	KEGG
Citrate cycle (TCA cycle)	7/20	5.4377E-11	10.265	1.7944E-9	3.0296E-10	0.30785	KEGG SMP
Ubiquinone and other terpenoid-quinone biosynthesis	1/9	3.8566E-9	8.4138	1.2341E-7	1.6712E-8	1.0	KEGG SMP
Phenylalanine, tyrosine and tryptophan biosynthesis	1/4	3.8566E-9	8.4138	1.2341E-7	1.6712E-8	0.0	KEGG SMP
Terpenoid backbone biosynthesis	1/18	6.6787E-9	8.1753	2.0036E-7	2.6047E-8	0.18571	KEGG
Pyruvate metabolism	4/22	8.5946E-9	8.0658	2.4924E-7	3.0472E-8	0.20684	KEGG SMP
Pentose phosphate pathway	9/22	1.8349E-8	7.7364	5.1377E-7	5.9634E-8	0.38831	KEGG SMP
Cysteine and methionine metabolism	2/33	2.1498E-8	7.6676	5.8045E-7	6.4494E-8	0.0	KEGG SMP SMP
Pentose and glucuronate interconversions	7/18	5.128E-8	7.2901	1.3333E-6	1.3755E-7	0.57812	KEGG
Glycolysis / Gluconeogenesis	5/26	5.2903E-8	7.2765	1.3333E-6	1.3755E-7	0.27796	KEGG SMP SMP
Tyrosine metabolism	4/42	6.3616E-8	7.1964	1.5268E-6	1.5506E-7	0.10816	KEGG SMP SMP
Glutathione metabolism	1/28	1.294E-7	6.8881	2.9761E-6	2.9685E-7	0.00709	KEGG SMP
Glycine, serine and threonine metabolism	4/33	6.4874E-7	6.1879	1.4272E-5	1.4056E-6	0.0697	KEGG SMP
Arginine and proline metabolism	2/38	8.6E-6	5.0655	1.806E-4	1.7653E-5	0.0	KEGG SMP
Amino sugar and nucleotide sugar metabolism	7/37	9.6305E-6	5.0164	1.9261E-4	1.8779E-5	0.19789	KEGG SMP SMP

Figure 9: Statistical results for the pathway analysis are interactive

15. Questions:
    a. What are the top 3 pathways that are predicted to be altered?

b. Which Pathway has the greatest proportion of metabolite matches (e.g. identified metabolites in that pathway)?
c. What is the name of the metabolite in the highest ranked pathway?

16. Select submit and then 'Generate Report'. Click on 'Analysis Report' and save. The report also provides an introduction to the principles of the pathways analysis and how it works. Links to various publications and tutorials can be found on the MetaboAnalyst website from which you can learn more about the functionality of the software and statistical tools.

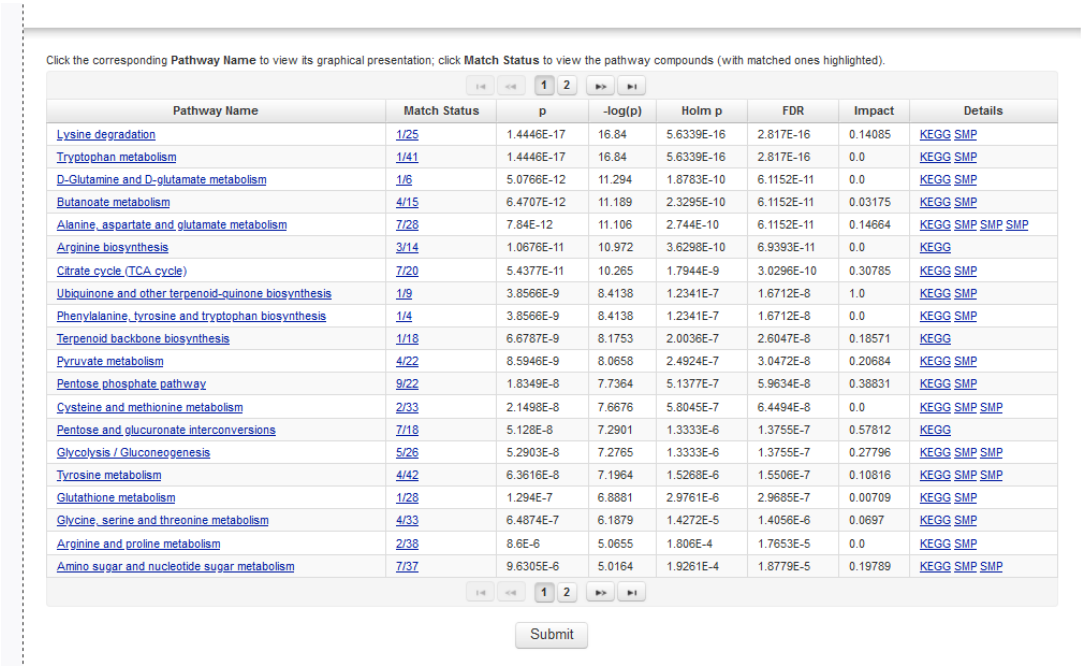


Figure 10: Download and save report for pathways analysis.

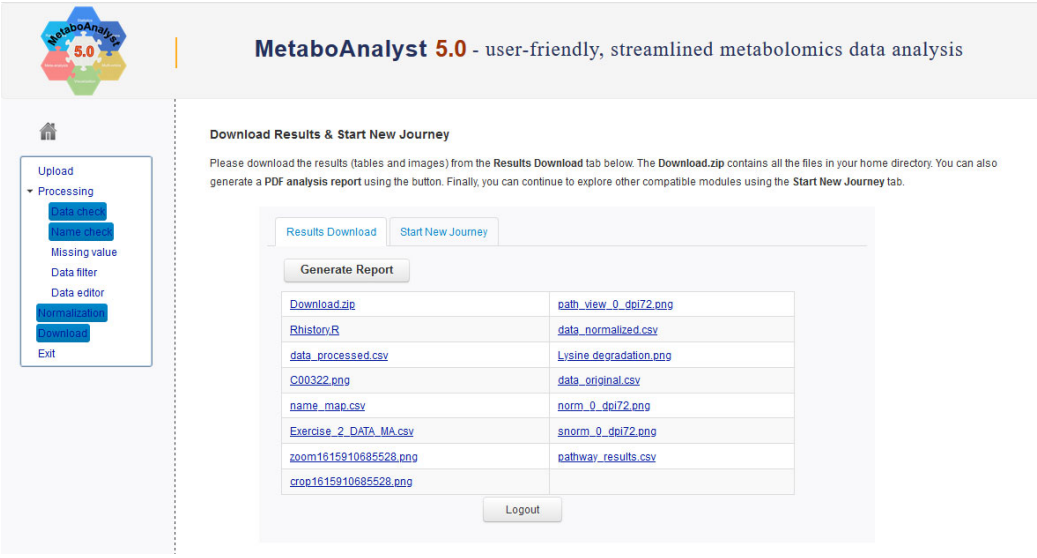


Figure 11: Generate a report

Part 2: Untargeted Pathways Analysis

This section provides a step by step guide to untargeted pathways analysis using '***Functional Analysis'*** in MetaboAnalyst (based on Mummichog algorithm).

1. Copy the .csv file called **'Exercise-3_untargeted.csv'** to your local computer (Note this is an entirely different dataset to the one used in Part 1).

2. Open MetaboAnalyst https://www.metaboanalyst.ca/

3. Select: **>>click here to start<<** which will open up the circular list of modules (see Figure 12 below).

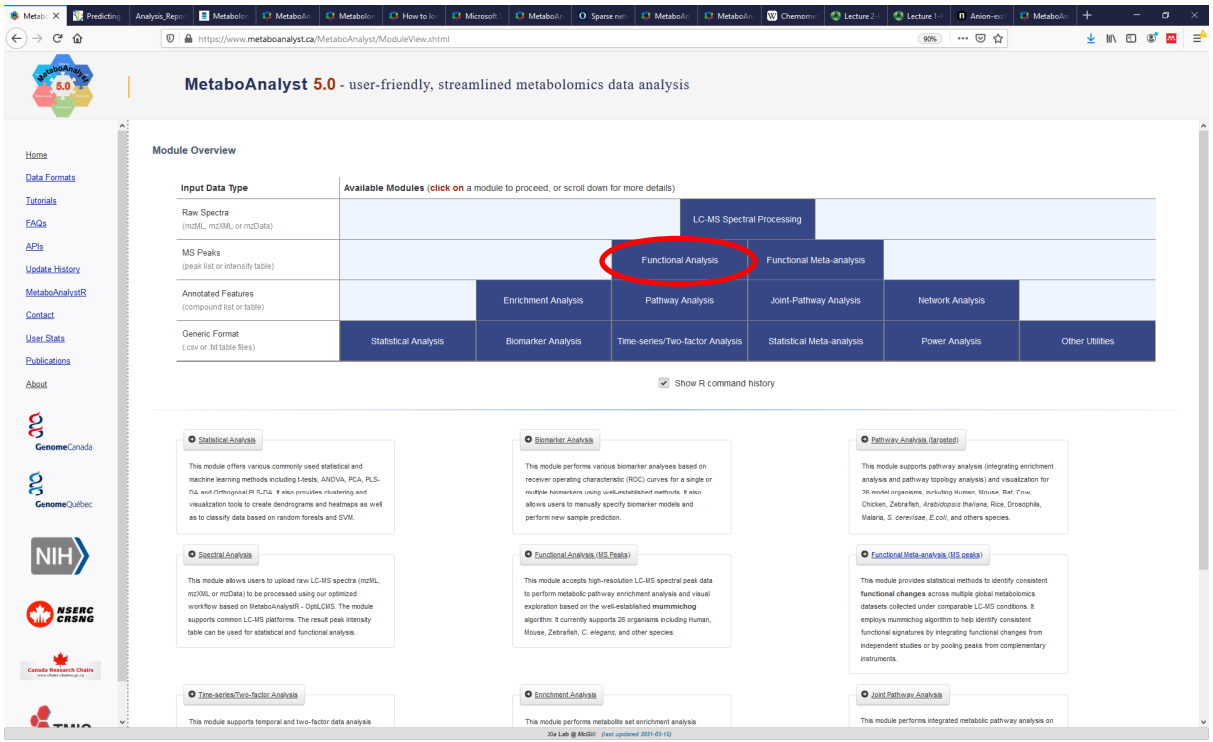


Figure 12: Pyramid list of modules in MetaboAnalyst.

4. **Select 'Functional Analysis'** tab **(see Figure 12)** which will open up a window (see **Figure 13** below). Select the '**Peak intensity Table' tab** at the top (note this is not the default) Please complete the details as shown in **Figure 13** (should be default). Select 'browse' and attach '**Exercise-3_untargeted.csv**' (available to download on the Teams site for the course).
5. **Press 'Proceed'**

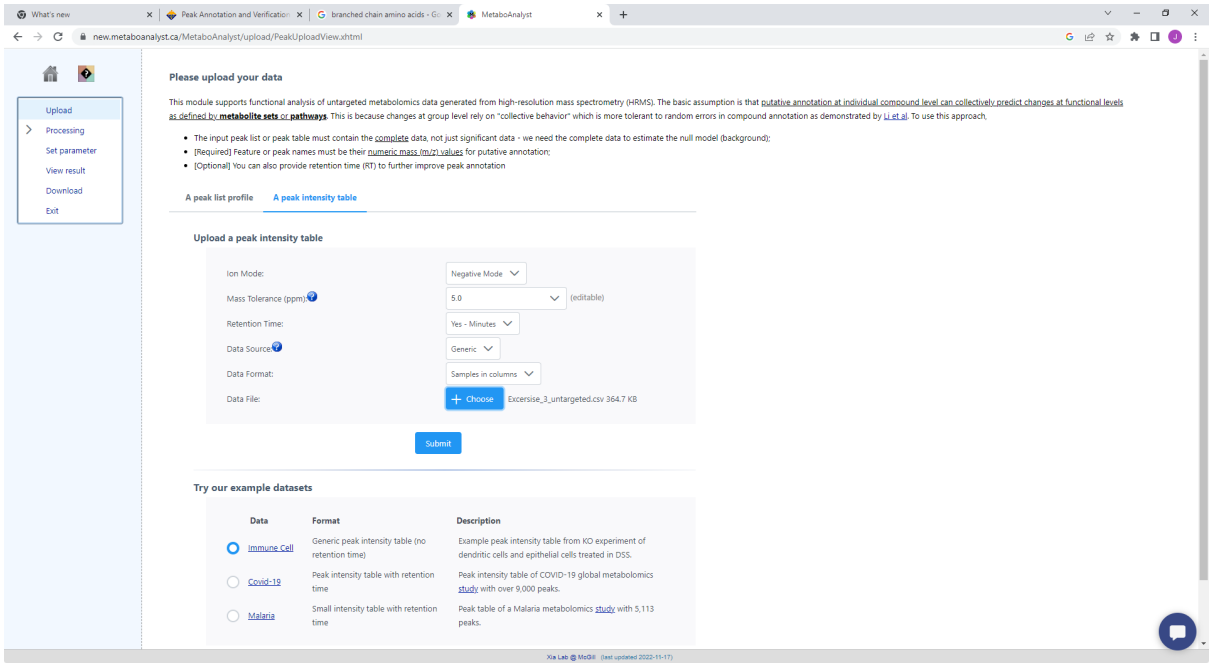


Figure 13: 'Upload your data' page (note that the 'peak intensity table' tab is selected at the top (not the default).

6. The **'Data Integrity Check'** page provides a summary of the data (Figure 14). Select **'Proceed'.**

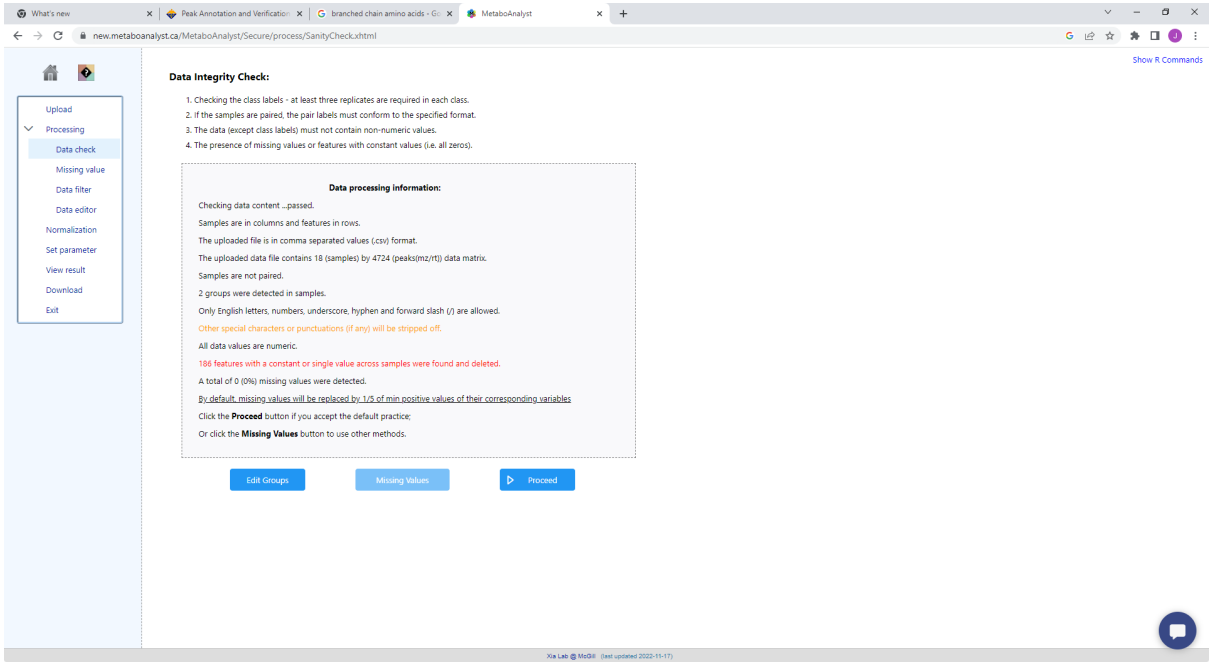


Figure 14: The 'Data Integrity Check' page (not it informs that some zero values were identified and removed)

7. The next page is for data filtering. We use the exact same parameters as for the statistical analysis e.g. Keep the default settings (IQR filtration). Click on submit and then ‘Proceed’.
8. The next page is the Normalisation overview we encountered during data processing for statistical analysis. You should the same settings determined as most appropaote for the statistical analysis. In this case:
    a. Normalisation by Sum
    b. Log Transformation
    c. Pareto Scaling

9. The next page provides parameters and setting for the pathways analysis. Note the options available on the ‘Set Parameters’ page (Figure 15). You can use the default settings for now. Scroll to the bottom and **click on ‘Proceed’**.

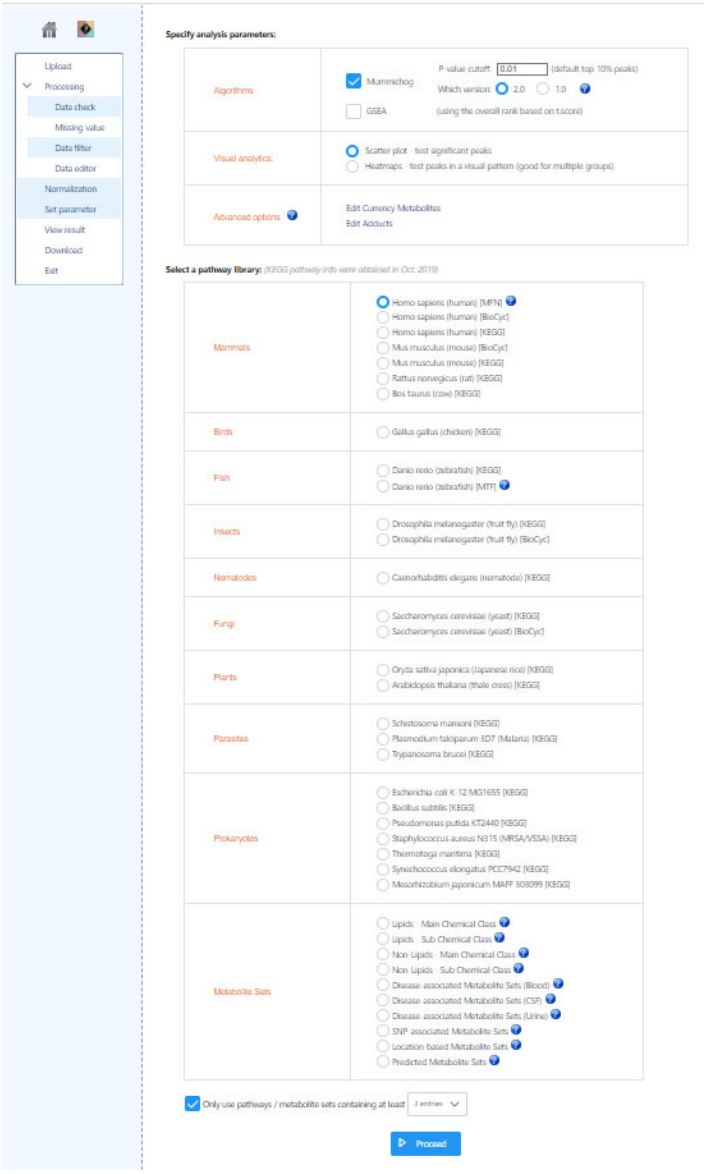


Figure 6: ‘Set Parameter’ page

10. The Mummichog '**Pathway Activity Profile'** is loaded (**Figure 16**). This looks very similar to the 'Pathways Analysis Results screen' shown for the targeted pathways analysis and works in a similar way. Hover over the un-annotated circled node (Figure 16). Glutathione metabolism is shown. The list of pathways and ranking is found below the figure ranked by significance (p-value). Note the data can be downloaded at this stage using the blue tabs above the pathway list. Either the Pathway Hits or Compound Hits (e.g. which have been putatively annotated). You can also click on the view link (under 'Details' to see which metabolites are in a pathway of interest and which have been identified and whether a particular metabolite hit is significant (red) or not significant (blue). See Figure 17 by way of example for the Glycine, Serine and Alanine pathway which is predicted to be the most significantly altered pathway. (**Figure 17**).

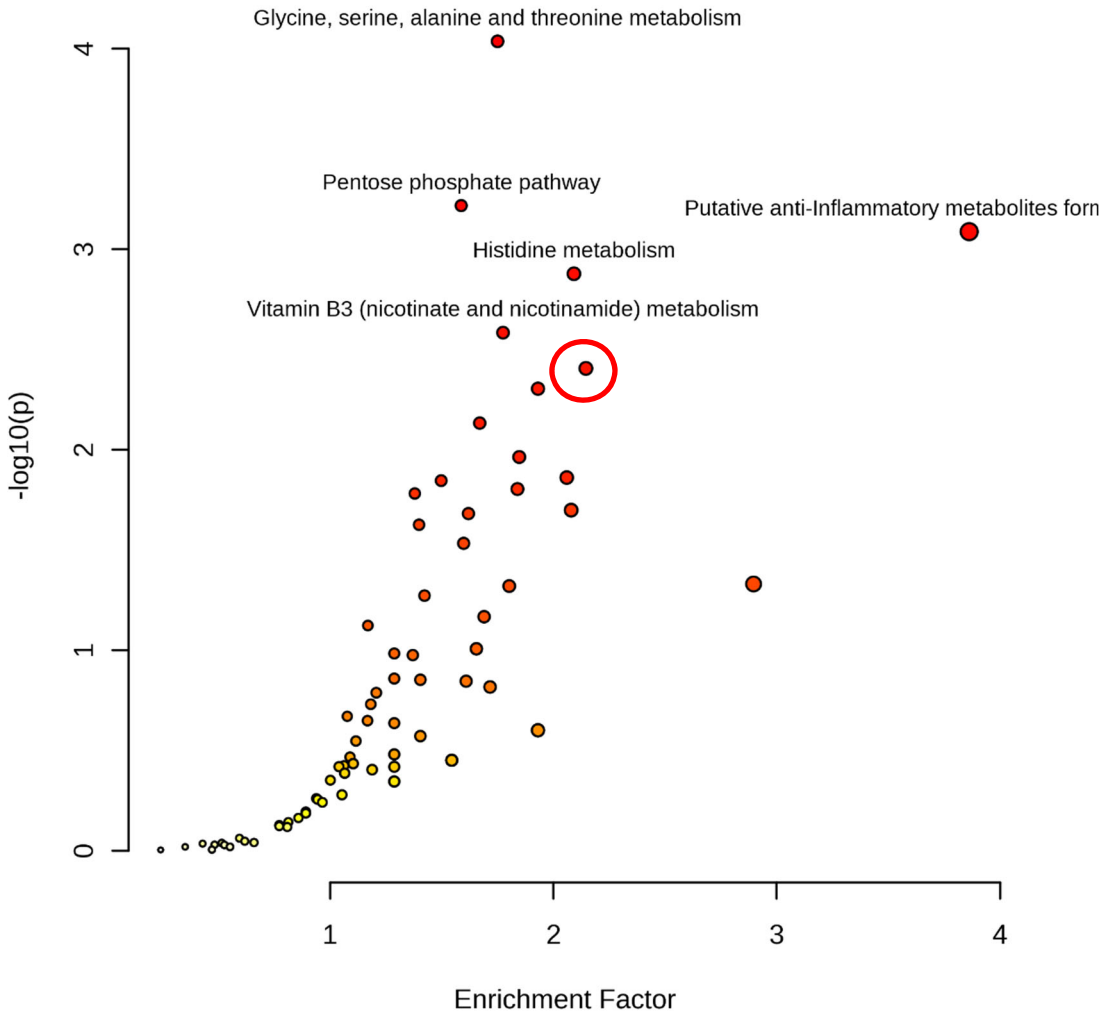


Figure 7: Mummichog Pathway Activity Profile.

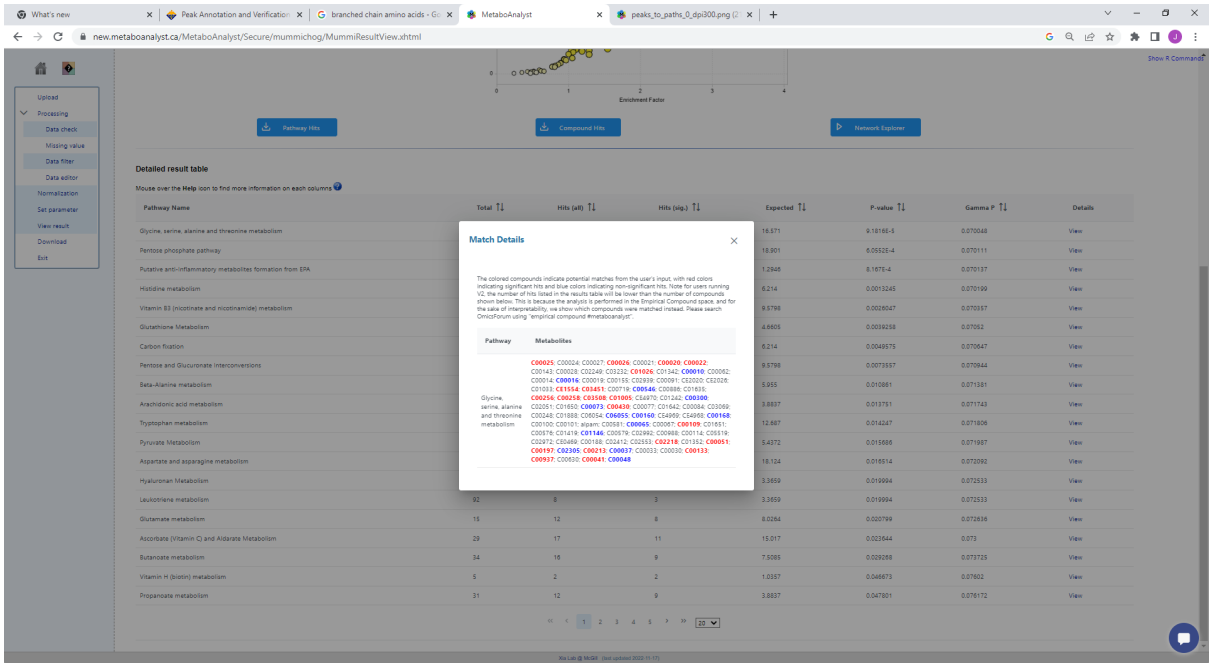


Figure 8: Which metabolite are identified can be viewed.

11. Select 'Network Explorer' on left hand side navigation panel. This opens a window showing the entire metabolic network template for the organism chosen on the **'Set Parameter'** page (Figure 18).

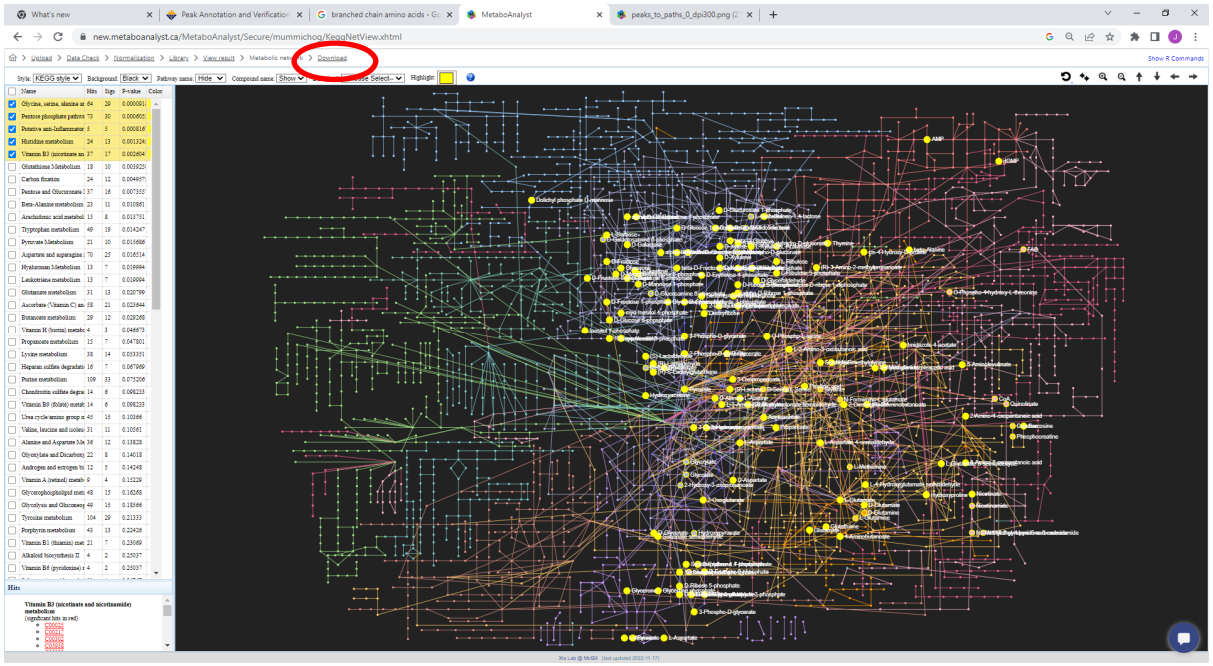


Figure 9: Metabolic Network page

12. Select the first five pathways via the tickbox next to 'Name' at the left hand side list of pathways, this will populate the pathways network with all matched compound

features for each pathway. Note the dropdown options at the top of the page which can be useful to configure the output.

13. Questions Part 2:

a. Which metabolic pathway has the highest proportion of putative identification via the untargeted pathways analysis in the top 6 pathways?

b. If you consider all pathways with a p-value <0.05 which broad areas of metabolism would you suggest are affected in the presence of IDH1 mutations?

c. Using the network explorer tab illustrate whether significantly altered pathways populate a similar or distinctly separate areas of metabolic space?

14. This completes the 'Functional Analysis Exercise. To download the MetaboAnalyst report click on the 'Download' at the top of the page (circled in Figure 18).

15. Select 'Generate Report' on the next page followed by 'Analysis Report' after it downloads (Figure 19).

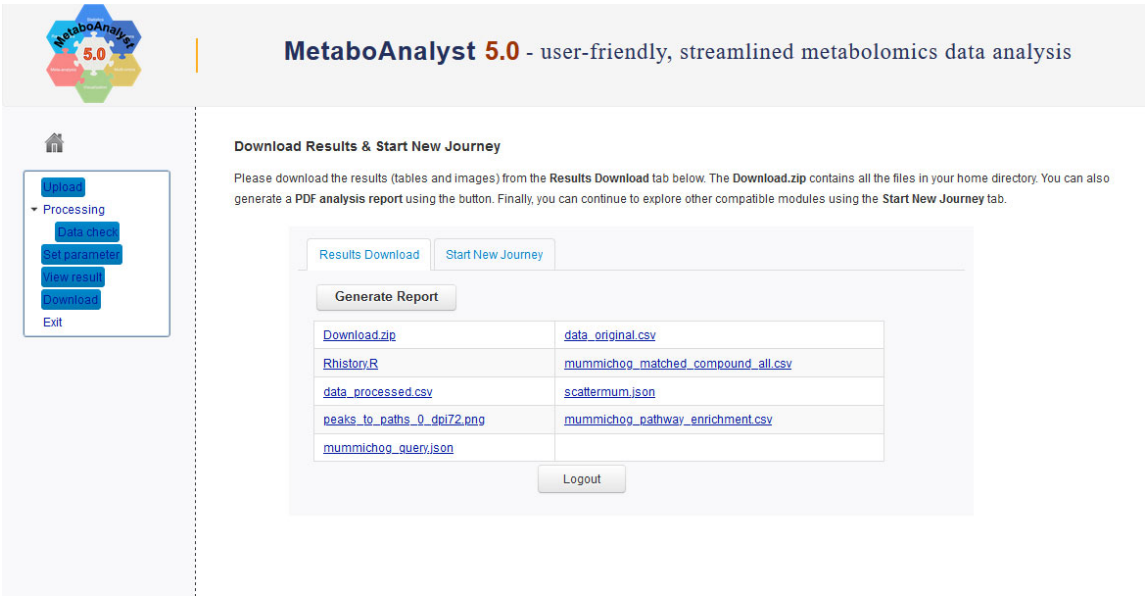


Figure 10: Click on 'analysis Report' to download the PDF report of MS Peaks to Pathways Analysis

16. A PDF report will open (**Figure 20**). Save this to your computer.

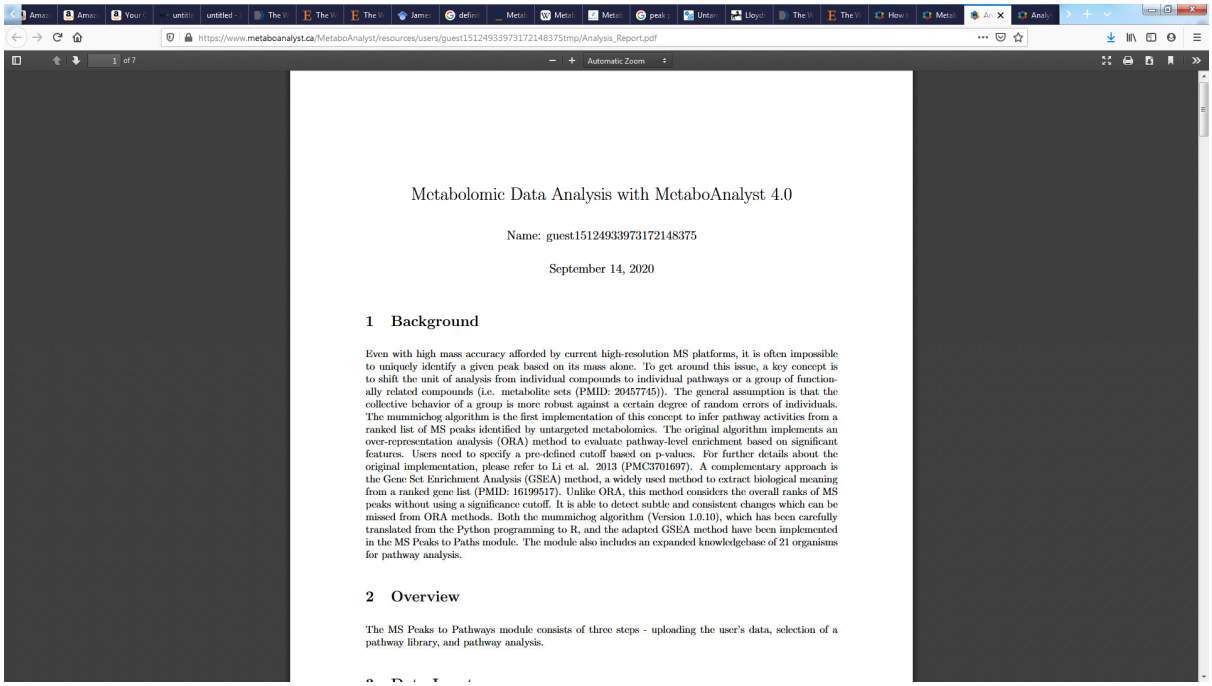


Figure 20: Save the PDF report for the MS Peaks to Pathways analysis.

Part 3: Multiomics

This section provides a step by step guide to multiomic pathways integration in the '***Joint Pathways Analysis'*** module in MetaboAnalyst.

17. Copy the .csv file called **'Exercise-3_untargeted.csv'** to your local computer (Note this is an entirely different dataset to the one used in Part 1).

18. Open MetaboAnalyst https://www.metaboanalyst.ca/

19. Select: **>>click here to start<<** which will open up the circular list of modules (see Figure 21 below).

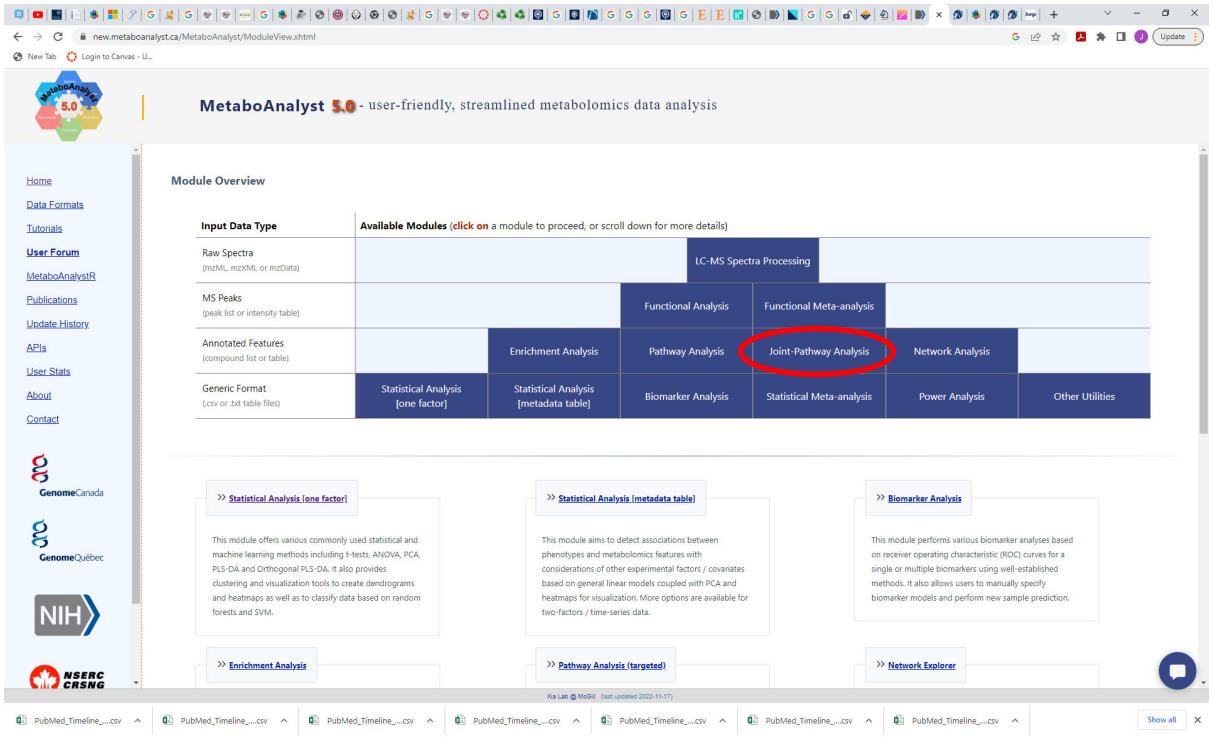


Figure 21: Pyramid list of modules in MetaboAnalyst.

20. **Select 'Joint-Pathways Analysis** tab **(see Figure 21)** which will open up a window. Open the '**Exercise_4_mulitomics.csv'** file. This contains column A and B with metabolite information and log2fold changes and Column D and E with transcriptomic information and corresponding log2fold change information (**Figure 22 below**). Copy and paste these data into the corresponding boxes in MetaboAnalyst (so shown in **Figure 23** below)
21. Make sure the organism is set to 'Homo sapiens (human)' and ID type is 'Official Gene Sumbol' for the transcriptomics data and for the Metabolomics data ensure 'Targeted (compound list)' is selected and ID-type is 'HMDB ID'.

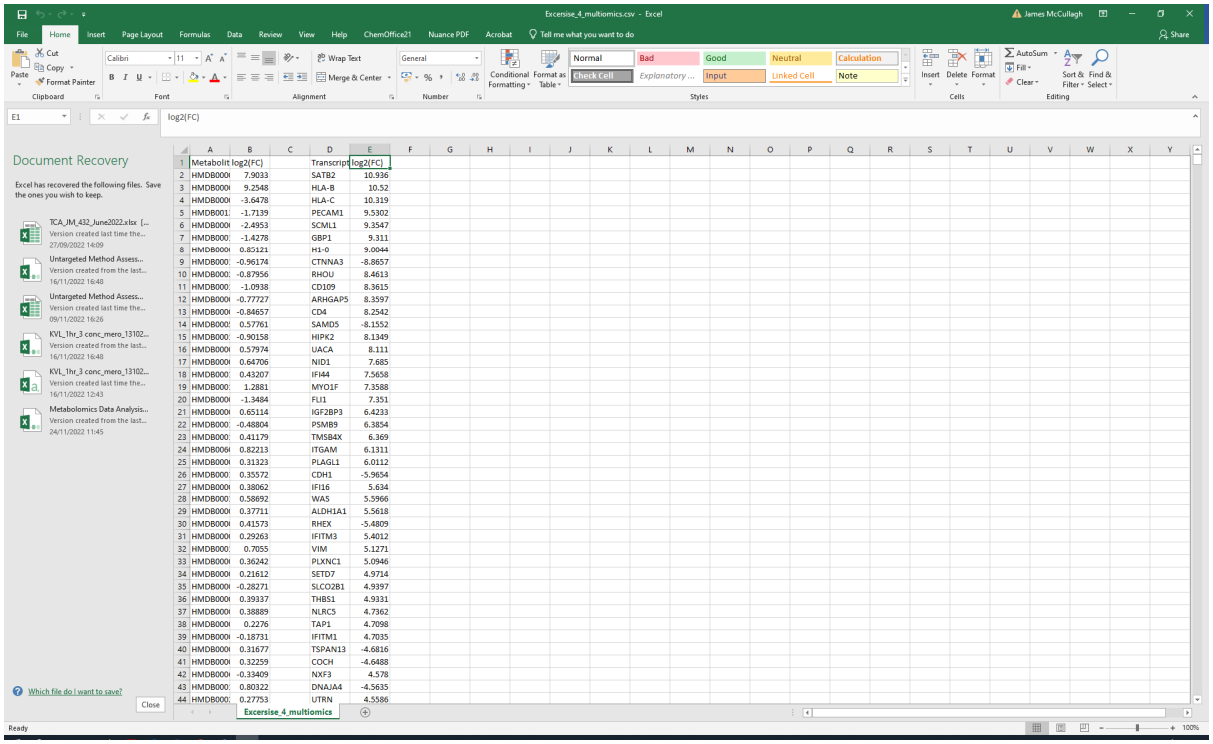


Figure 22: Data found in Exercise_4_mulitomics.csv'

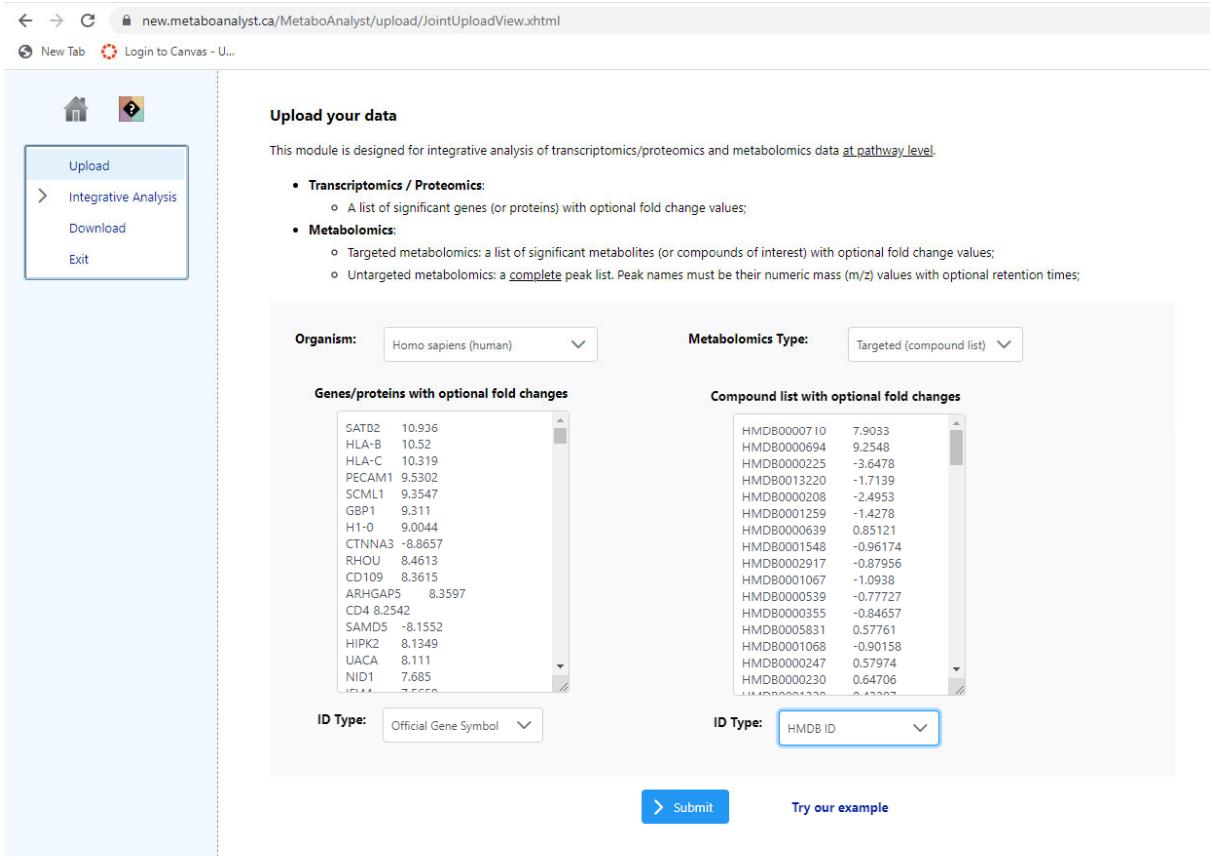


Figure 23: Upload the metabolomics and transcriptomics data.

22. Click on **'Submit'**
23. Information about the gene and metabolite name mapping is given and the opportunity to exclude either (**Figure 24**). Scroll to the bottom and select 'Proceed'

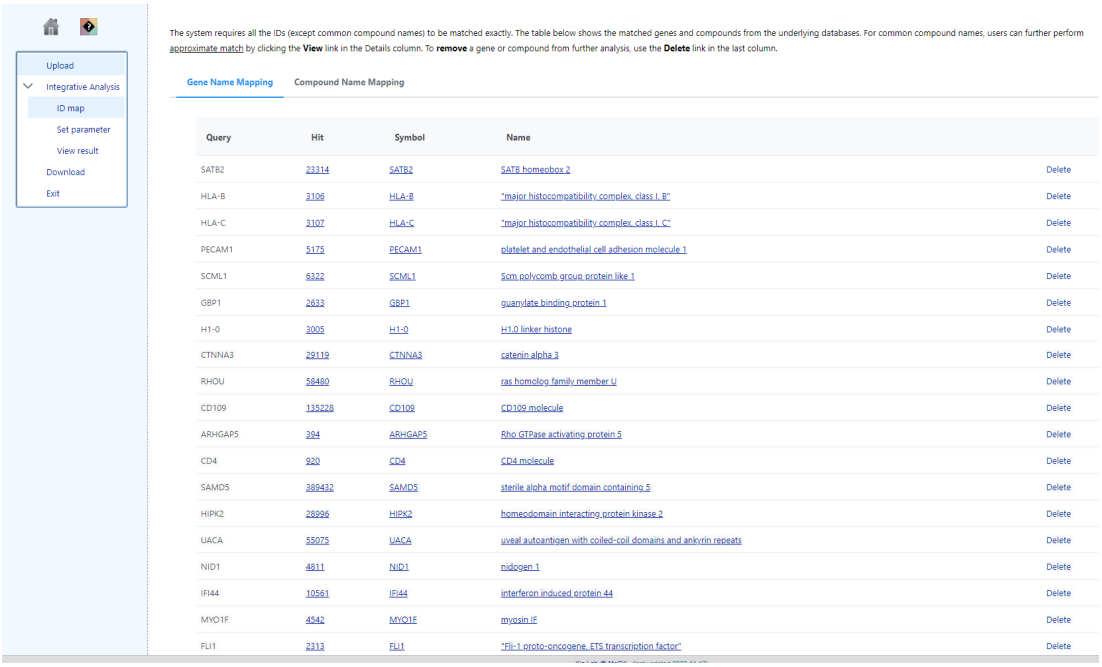


Figure 24: The 'Data Integrity Check' page (not it informs that some zero values were identified and removed)

24. The **Parameter Setting Page** enables you to choose how the data will be integrated and the type of algorithm used for data integration (**Figure 25**). For the purposes of this example we will keep all the default setting. Scroll to the bottom of the page and **click on 'Proceed'.**

25. The Results View (**Figure 25**) provides the results of the multiomics integration in the form of a Pathways 2-D plot familiar from the Pathways Analysis modules. This now has combined metabolomics and transcriptomics data. you can investigate which pathways shows greatest significance and impact by hovering over the red points in the top right of the 2-D plot. Click on the most significant and highest impact pathway (circled in red) and wait a few seconds. The pathway (KEGG format) is provided with the genes (square) and metabolites (circles) colour coded according to whether they are significantly altered in abundance (red) or not (green). If you hover over a gene or metabolite further information is provided (**Figure 25**).

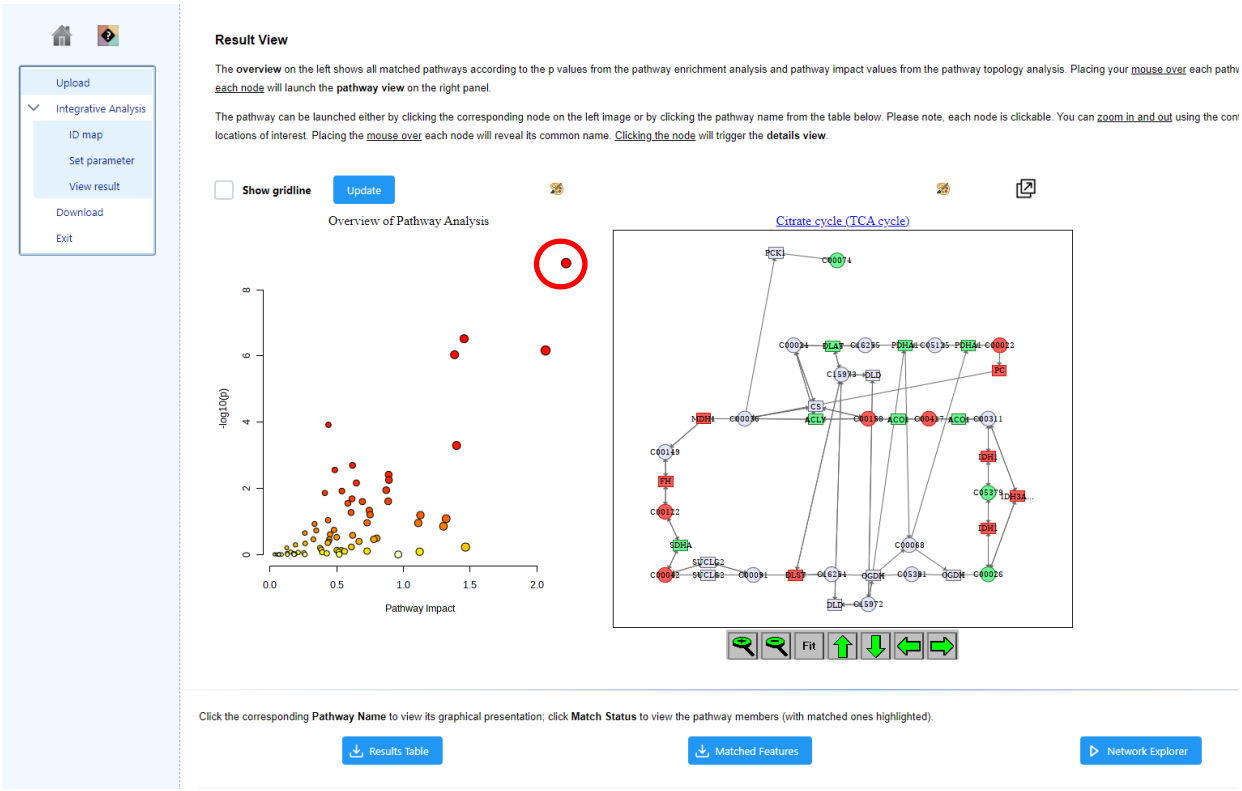
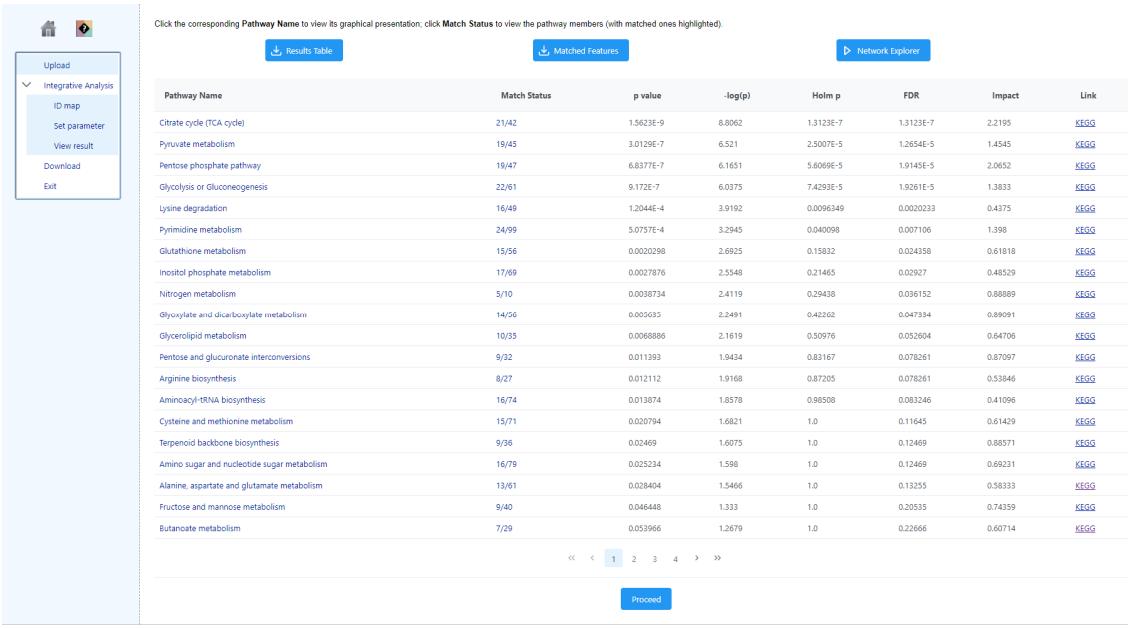


Figure 25: Results from multi-omic integration

26. If you scroll down you can see the results tables with associated statistics. Note there is a FDR-corrected significance value associated with each pathway (**Figure 26**). You can download further information from the analysis using the blue boxes provided.



Pathway Name	Match Status	p value	-log(p)	Holm p	FDR	Impact	Link
Citrate cycle (TCA cycle)	21/42	1.5623E-9	8.8062	1.3123E-7	1.3123E-7	2.2195	KEGG
Pyruvate metabolism	19/45	3.0129E-7	6.521	2.5007E-5	1.2654E-5	1.4545	KEGG
Pentose phosphate pathway	19/47	6.8377E-7	6.1651	5.6069E-5	1.9145E-5	2.0652	KEGG
Glycolysis or Gluconeogenesis	22/61	9.172E-7	6.0375	7.4293E-5	1.9261E-5	1.3833	KEGG
Lysine degradation	16/49	1.2044E-4	3.9192	0.0096349	0.0020233	0.4375	KEGG
Pyrimidine metabolism	24/99	5.0757E-4	3.2945	0.040098	0.007106	1.398	KEGG
Glutathione metabolism	15/56	0.0020298	2.6925	0.15832	0.024358	0.61818	KEGG
Inositol phosphate metabolism	17/69	0.0027876	2.5548	0.21465	0.02927	0.48529	KEGG
Nitrogen metabolism	5/10	0.0038734	2.4119	0.29438	0.036152	0.88889	KEGG
Glyoxylate and dicarboxylate metabolism	14/56	0.005035	2.2491	0.42262	0.047334	0.89091	KEGG
Glycerolipid metabolism	10/35	0.0068886	2.1619	0.50976	0.052604	0.64706	KEGG
Pentose and glucuronate interconversions	9/32	0.011393	1.9434	0.83167	0.078261	0.87097	KEGG
Arginine biosynthesis	8/27	0.012112	1.9168	0.87205	0.078261	0.53846	KEGG
Aminoacyl-tRNA biosynthesis	16/74	0.013874	1.8578	0.98508	0.083246	0.41096	KEGG
Cysteine and methionine metabolism	15/71	0.020794	1.6821	1.0	0.11645	0.61429	KEGG
Terpenoid backbone biosynthesis	9/36	0.02469	1.6075	1.0	0.12469	0.88571	KEGG
Amino sugar and nucleotide sugar metabolism	16/79	0.025234	1.598	1.0	0.12469	0.69231	KEGG
Alanine, aspartate and glutamate metabolism	13/61	0.028404	1.5466	1.0	0.13255	0.58333	KEGG
Fructose and mannose metabolism	9/40	0.046448	1.333	1.0	0.20535	0.74359	KEGG
Butanoate metabolism	7/29	0.053966	1.2679	1.0	0.22666	0.60714	KEGG

Figure 11: List of pathways predicted by multiomic integration and associated statistics.

27. Click on 'Proceed. To be taken to the 'Download Results & Start New Journey' screen select 'Generate Report' and download this to your computer.

Questions Part 3: Multiomics

a. How many pathways were predicted with an FDR-corrected p-value <0.05?

b. Which metabolic process appears to be most significantly altered?

c. Click on the KEGG link for the 'Citrate cycle (TCA cycle)' in the results table: Why are some of the enzymes labelled in green and some left white?

d. Predict what will happen to the abundance of the metabolite 1) precursor and 2) product if the corresponding metabolic enzyme's gene is upregulated (assume the transcriptional change leads to a change in the corresponding active enzyme)?

e. How might a metabolite's abundance be altered independently from direct transcriptional changes?

THIS COMPLETES THE STEP BY STEP GUIDE TO PATHWAYS ANALYSIS

LIST OF TASKS FOR PATHWAYS ANALYSIS (NON-STEP BY STEP)

Part 1: Targeted pathways analysis

- Upload the '**Exercise_2_DATA_MA.csv'** dataset in the Pathways Analysis module. Use *Homo sapiens* as the organism metabolic pathway library.
- Identify the pathway which has the most statistically significant changes. Identify the pathway having the greatest metabolic impact.
- Download the data and generate a report. Select 'Analysis Report' and save it.

Questions:

a. What are the top 3 pathways that are predicted to be altered?
   - **Answer:**

b. Which Pathway has the greatest proportion of metabolite matches (e.g. identified metabolites in that pathway)?
   - **Answer:**

c. What is the name of the metabolite in the highest ranked pathway?
   - **Answer:**

Part 2: Functional analysis of an untargeted dataset

- Open the 'Functional Analysis' module and load **'Exercise-3_untargeted.csv'** datafile using the following parameters (negative ion mode; 5ppm mass tolerance; retention time not present; ranked by p-values.
- Create a metabolic network model from the untargeted dataset (using *Homo sapiens* as the organism metabolic pathway library).
- Identify the most significantly altered metabolic pathway
- Identify the pathway showing the greatest metabolic impact.
- Explore these pathways using the metabolic network visualisation tool.
- Download the data and generate a report. Select the Analysis Report and save it.

Questions Part 2:

d. Which metabolic pathway has the highest proportion of putative identification via the untargeted pathways analysis in the top 6 pathways?

e. If you consider all pathways with a p-value <0.05 which broad areas of metabolism would you suggest are affected in the presence of IDH1 mutations?

f. Using the network explorer tab illustrate whether significantly altered pathways populate a similar or distinctly separate areas of metabolic space?

Part 3: Multiomics

- Use the Joint Pathways Analysis module in MetaboAnalyst 5.0 to investigated the transcriptome and metabolome data in **'Exercise_4_mulitomics.csv'.**
- Create a pathway integration map using default settings.
- Identifying the pathways that are predicted to be significantly altered.
- Answer the questions on Part 3 below.

Questions Part 3: Multiomics

g. How many pathways were predicted with an FDR-corrected p-value <0.05?

h. Which metabolic process appears to be most significantly altered?

i. Click on the KEGG link for the 'Citrate cycle (TCA cycle)' in the results table: Why are some of the enzymes labelled in green and some left white?

j. Predict what will happen to the abundance of the metabolite 1) *precursor* and 2) *product* if the corresponding metabolic enzyme's gene is upregulated (assume the transcriptional change leads to a change in the corresponding active enzyme)?

k. How might a metabolite's abundance be altered independently from direct transcriptional changes?