# Part 2 (Exercises 2,3 and 4):

# Pathways Analysis and Multiomics



# 3<sup>rd</sup> 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

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

	Module Overview						
its	Input Data Type	Available Modules (click on a	module to proceed, or scroll dowr	for more details)			
	Raw Spectra			LC-MS Spectra	al Processing		
	(mzML, mzXML or mzData)						
ory	MS Peaks (peak list or intensity table)			Functional Analysis	Functional Meta-analysis		
lystR	Annotated Features		Enrichment Analysis	Pathway Analysis	Joint-Pathway Analysis	Network Analysis	
	Generic Format						
	(.csv or .bd table files)	Statistical Analysis	Biomarker Analysis	Time-seriou Analysis	Statistical Meta-analysis	Power Analysis	Other Utili
				Show R command h	history		
	<ul> <li>Statistical Analysis</li> </ul>		O Biomarker Analysis		O Pati	hway Analysis (bargeted)	
Canada	This module offers various commonly used s	tatistical and	This module performs varie	ous biomarker analyses based on	This m	odule supports pathway analysis (integrating	enrichment
	machine learning methods including t-tests, A DA and Orthogonal PLS-DA. It also provides	NOVA, PCA, PLS-	receiver operating charact multiple biomarkers using v	eristic (ROC) curves for a single or vell-established methods. It also	analys 26 mor	is and pathway topology analysis) and visuals del organisms, including Human, Mouse, Rat, C	ization for
Québec	visualization tools to create dendrograms and as to classify data based on random forests	d heatmaps as well and SVM.	allows users to manually a perform new sample predi	pecify biomarker models and ction.	Chicke Malarie	n, Zebrafish, Arabidops/s thaliana, Rice, Dros a, S. cerevisae, E.coli, and others species.	sophila,
<b>\</b>	Spectral Analysis		Functional Analysis (HS	Peaks)	O Fur	ictional Meta-analysis (MS, peaks)	
/	This module allows users to upload raw LC-M	IS spectra (mzML,	This module accepts high-r	esolution LC-MS spectral peak data	This ma	odule provides statistical methods to identify or	onsistent
	mzXML or mzData) to be processed using ou	r optimized MS. The module	to perform metabolic pathw exploration based on the w	ay enrichment analysis and visual ell-established mummichog	function	onal changes across multiple global metabolo ts collected under comparable LC-MS condition	omics ns. t
ERC	workflow based on MetaboAnalystR - OptiLC						
SERC ISNG	workflow based on MetaboAnalystR - OptLC supports common LC-MS platforms. The resu table can be used for statistical and functions	it peak intensity il analysis.	algorithm. It currently suppr Mouse, Zebrafish, C. elege	orts 26 organisms including Human, ins, and other species.	employ functio	is mummichog algorithm to help identify consist nal signatures by integrating functional change	es from

- 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.

5.0	MetaboAnalyst 5.0 - user-friendly, streamlined metabolomics data analysis
fit	Please upload your data use one of the options below Compound List Concentration Table Metabolomics Workbench Data
▶ Pathway Download Exit	Upload your concentration data (.csv or .txt) Group Label: Discrete (Classification) Continuous (Regression) ID Type:Please specify · Data Format: Samples in rows · Data File: Browse No file selected.
	Use the example data Data Description Uninary metabolite concentrations from 77 cancer patients measured by 1H NMR. Phenotype: N - cachexic; Y - control Submit

Figure 2: Select 'concentration table'.

staboAnage 5.0	MetaboAnalyst 5.0 - user-friendly, streamlined metabolomics data analysis
ñ	Please upload your data use one of the options below
Extrast Processing Pathway Download Exit	Compound List       Concentration Table       Metabolomics Workbench Data         Upload your concentration data (.csv or .txt)       Group Label:       Discrete (Classification)       Continuous (Regression)         ID Type:       HHDB D       Image: Continuous (Regression)       Image: Continuous (Regression)         Data Format:       Samples in columns       Image: Continuous (Regression)         Data File:       Browse       Exercise_2_DATA_MA.csv         Image: Use the example data       Image: Concentrations from 77 cancer patients measured by 1H NIMR. Phenotype: N-Cachexic; Y - control         Submit       Submit

**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).

â	Data Integrity Check:
	<ol> <li>Checking the class labels - at least three replicates are required in each class.</li> </ol>
Upload	<ol><li>If the samples are paired, the pair labels must conform to the specified format.</li></ol>
Processing	3. The data (except class labels) must not contain non-numeric values.
Data check	<ol> <li>Ine presence or missing values or reatures with constant values (i.e. all zeros).</li> </ol>
Name check	
Missing value	Data processing information:
Data inter	Checking data contentpassed.
Normalization	Samples are in columns and features in rows.
Download	The uploaded file is in comma separated values (.csv) format.
Evit	The uploaded data file contains 18 (samples) by 109 (compounds) data matrix.
EAR	Samples are not paired.
	2 groups were detected in samples.
	Only English letters, numbers, underscore, hyphen and forward slash (/) are allowed.
	Other special characters or punctuations (if any) will be stripped off.
	All data values are numeric.
	1 features with a constant or single value across samples were found and deleted.
	A total of 0 (0%) missing values were detected
	By default mission values will be replaced by 1/5 of min positive values of their corresponding variables
	Click the Skin button if you accent the default practice:
	Once the step balanting the active practice practice of the step o
	Or circk title winssing value imputation to use other methods.

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.

-						
	HMDB0000267	Pyroglutamic acid	HMDB0000267	7405	<u>C01879</u>	
	HMDB0000243	Pyruvic acid	HMDB0000243	1060	<u>C00022</u>	
d	HMDB0003072	Quinic acid	HMDB0003072	6508	<u>C00296</u>	
ssing	HMDB0000232	Quinolinic acid	HMDB0000232	<u>1066</u>	<u>C03722</u>	
a check	HMDB0003213	Raffinose	HMDB0003213	<u>10542</u>	<u>C00492</u>	
me check	HMDB0001548	D-Ribose 5-phosphate	HMDB0001548	<u>439167</u>	<u>C03736</u>	
a filter	HMDB0000618	D-Ribulose 5-phosphate	HMDB0000618	439184	<u>C00199</u>	
ta editor	HMDB0000792	Sebacic acid	HMDB0000792	<u>5192</u>	<u>C08277</u>	
alization	HMDB0060274	Sedoheptulose 1,7-bisphosphate	HMDB0060274	164735	<u>C00447</u>	
load	HMDB0060509		<i></i>	-	-	View
	HMDB0001068	D-Sedoheptulose 7-phosphate	HMDB0001068	22833559	<u>C05382</u>	
	HMDB0000247	Sorbitol	HMDB0000247	5780	<u>C00794</u>	
	HMDB0005831	Sorbitol-6-phosphate	HMDB0005831	618	C02810	
	HMDB0000254	Succinic acid	HMDB0000254	1110	<u>C00042</u>	
	HMDB0001259	Succinic acid semialdehyde	HMDB0001259	1112	<u>C00232</u>	
	HMDB0001227	5-Thymidylic acid	HMDB0001227	<u>9700</u>	<u>C00364</u>	
	HMDB0001342	Thymidine 5'-triphosphate	HMDB0001342	64968	<u>C00459</u>	
	HMDB0000935	Uridine diphosphate glucuronic acid	HMDB0000935	17473	C00167	
	HMDB0000286	Uridine diphosphate glucose	HMDB0000286	<u>53477679</u>	<u>C00029</u>	
	HMDB0000300	Uracil	HMDB0000300	<u>1174</u>	<u>C00106</u>	
	HMDB0000289	Uric acid	HMDB0000289	1175	<u>C00366</u>	
	HMDB0000296	Uridine	HMDB0000296	6029	<u>C00299</u>	
	HMDB0000288	Uridine 5'-monophosphate	HMDB0000288	<u>6030</u>	<u>C00105</u>	
	HMDB0000290	Uridine diphosphate-N-acetylglucosamine	HMDB0000290	9547196	<u>C00043</u>	
	HMDB0000285	Uridine triphosphate	HMDB0000285	6133	<u>C00075</u>	
	HMDB0001554	Xanthylic acid	HMDB0001554	73323	<u>C00655</u>	
	HMDB0002917	D-Xylitol	HMDB0002917	<u>6912</u>	<u>C00379</u>	
	HMDB0000868	Xylulose 5-phosphate	HMDB0000868	439190	<u>C00231</u>	
	HMDB0001487	NADH	HMDB0001487	928	<u>C00004</u>	
	HMDB0000295	Uridine 5'-diphosphate	HMDB0000295	<u>6031</u>	C00015	

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)

	Normalization overview:
oad cessing Data check	The normalization procedures are grouped into three categories. The sample normalization allows general-purpose adjustment for differences among your sample, data transformation and scaling are two different approaches to make individual features more comparable. You can use one or combine them to achieve better results.
Aime Check Aissing value Data filter Data editor	Sample Normalization
malization vnload	Sample-specific normalization (i.e. weight, volume) Specify  Normalization by sum
	Normalization by median     Normalization by reference sample (PQN)     Specify
	Normalization by a pooled sample from group         Specify           Normalization by reference feature         Specify
	Quantile normalization
	Cube root transformation (generalized logarithm transformation or glog)     Cube root transformation (lakes the cube root of data values)
	Data scaling
	None     Mean centering (mean-centered only)
	Auto scaling (mean-centered and divided by the standard deviation of each variable)     The scaling (mean-centered and divided by the square root of the standard deviation of each variable)
	Range scaling (mean-centered and divided by the range of each variable)

**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'

ñ	Specify pathway analysis parameter	'S:	
pload rocessing	Visualization method	Scatter plot (testing significant features)     Heatmaps (testing your selected features)	
Data check Name check	Enrichment method	<ul> <li>Global Test</li> <li>Global Ancova</li> </ul>	
Data filter Data editor	Topology analysis	Relative-betweeness Centrality     Out-degree Centrality	
ormalization ownload xit	Reference metabolome	Use all compounds in the selected pathway library     Upload your own reference metabolome	
	Select a pathway library: (KEGG pathw	ay info were obtained in Oct. 2019)	
	Mammais	Homo sapiens (KEGG)     Homo sapiens (SMPDB)     Mus musculus (KEGG)     Mus musculus (SMPDB)     Ratlus norvegicus (rat) (KEGG)     Bos taurus (cow) (KEGG)	
	Birds	Gallus gallus (chicken) (KEGG)	
	Fish	Danio rerio (zebrafish) (KEGG)	
	Insects	Drosophila melanogaster (fruit fly) (KEGG)	
	Nematodes	Caenorhabditis elegans (nematode) (KEGG)	
	Fungi	Saccharomyces cerevisiae (yeast) (KEGG)	
	Plants	Oryza sativa japonica (Japanese rice) (KEGG)     Arabidopsis thaliana (thale cress) (KEGG)     Chorella variabilis (green alga) (KEGG)	
	Parasites	Schistosoma mansoni (KEGG) Plasmodium faiciparum 307 (Malaria) (KEGG) Plasmodium vivax (Malaria) (KEGG)	

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 - <u>light blue</u> means those metabolites are not in your data and are used as background for enrichment analysis; <u>grey</u> 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.]

8

setaboAnays.	MetaboAnalyst 5.0 - user-friendly	r, streamlined metabolomics c	lata analysis
Upload - Processing Data check Massing value Data filter Data editor Mormatication Download	Result View: The metabolome view on the left shows all matched pathways according to the put pathway topology analysis. Placing your <u>mouse over</u> each pathway node will reveal The pathway can be launched either by clicking the corresponding node on the left (compound) is clickable. You can zoom in and out using the control buttons below, metabolite node will reveal its common name. <u>Clicking the node</u> will trigger the corr About compound colors within the pathway - <u>light blue</u> means those metabolites an metabolite is not in your data and is also excluded from enrichment analysis (only yellow to red) means the metabolites are in the data with different levels of significant of the second sec	values from the pathway enrichment analysis and path its pathway name. <u>Clicking each node</u> will launch the image or by clicking the pathway name from the table and then <u>drag</u> the image to the locations of interest. Pi <b>mpound</b> view of the selected compound. re not in your data and are used as background for en applicable if you have uploaded a custom metabolom ance.	way impact values from the pathway view on the right panel. below. Please note, each node acing the <u>mouse over</u> each ichment analysis; <u>ore</u> y means the profile); other colors (varying from
Exit	Show gridline Update Overview of Pathway Analysis	Ubiquinone and other terpenoid-quit	CO3313 CO5849 CO2059

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	р	-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 SM
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.

	14	<1 2	14 44				
Pathway Name	Match Status	р	-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	<u>1/41</u>	1.4446E-17	16.84	5.6339E-16	2.817E-16	0.0	KEGG SMP
D-Glutamine and D-glutamate metabolism	<u>1/6</u>	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	<u>3/14</u>	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	<u>1/9</u>	3.8566E-9	8.4138	1.2341E-7	1.6712E-8	1.0	KEGG SMP
Phenylalanine, tyrosine and tryptophan biosynthesis	<u>1/4</u>	3.8566E-9	8.4138	1.2341E-7	1.6712E-8	0.0	KEGG SMP
Terpenoid backbone biosynthesis	<u>1/18</u>	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	<u>9/22</u>	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
<u>Glycolysis / Gluconeogenesis</u>	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	<u>4/33</u>	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 10: Download and save report for pathways analysis.

staboAnays 5.0	Metabo	Analyst <mark>5.0</mark> - user	-friendly, streamlined m	etabolomics data analysis
Upload • Processing Data check (some check) Missing value Data filter Data editor Hormatization Oxwinaat Exit	Download Results & Please download the resul generate a PDF analysis ro Results DD Generat Download Rhistor/R data_proc C00322.p name_ma Exercise zoom1515	Start New Journey Its (tables and images) from the Result sport using the button. Finally, you can winload Start New Journey the Report  20  20  20  20  20  20  20  20  20  2	Its Download tab below. The Download.zip of continue to explore other compatible module continue to explore other compatible module data.org         path_view_0_doir2.eng         data_normalized.csy         Lysine_degradation.ong         data_org/onal.csy         norm_0_doir2.eng         androw_degradation.csy         norm_0_doir2.eng         pathway_results.csy	ontains all the files in your home directory. You can also is using the Start New Journey tab.
		Lo	gout	

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/
- 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'

۲	What's new		×   🔶 Pe	ak Annotation and Verification	× G branched chain amino acids - Go	× 😣 MetaboAnalyst	× +				~	-	ð	×
÷	→ C  mew	v.metabo	oanalyst.ca/M	letaboAnalyst/upload/Peak	kUploadView.xhtml					GÉ	\$	*	•	:
	Processing     Steparameter     Vew result     Download     Evit		Plea This r as de	se upload your data nodule supports functional al time is un retablic sets or The input yeak its or peak (optional) you can also pro tipeurier() farture or peak (optional) you can also pro tipeurier() farture or tipeurier() farture or ti	Inalysis of untargeted metabolomics data g attbouys This is because changes at groc table must contain the <u>complete</u> data, not manner must be the <u>unment mass must must</u> must not a the <u>unment mass must</u> vide retention time (RT) to further improve ak intensity table ty table	penerated from high-resolution mass spect p level rely on "collective behavior" which just significant data - we need the comple diales for putties annotation: peak annotation Negative Mode 50 Vie - Minutes Genetic Samples in sclumes + Choose Excessive Juntargeted.	trometry (HRMS). The basic is more tolerant to random ete data to estimate the null oley	assumption is that <u>outative annotation at individual</u> errors in compound annotation as demonstrated by model (background):	somoound level can sollectively predict changes. Li <u>t et al</u> . To use this approach,	at function	levels			
		Try our example datasets Data Format		Subr asets Format	nt Description									
				O Immune Ce	Generic peak intensity table (no retention time)	Example peak intensity table from KO dendritic cells and epithelial cells treat	experiment of ted in DSS.							
				Covid-19	time	study with over 9,000 peaks.	ar metabolomics							1
				O <u>Malaria</u>	Small intensity table with retention time	Peak table of a Malaria metabolomics peaks.	study with 5,113						C	).
						Xia Lab @ McGil	I (last updated 2022-11-17)							

**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'.** 

What's new	x   🔶 Peak Annotation and Verification x   G branched chain amino acids - G x 🔅 MetaboAnalyst x +	✓ - Ø ×
$\leftrightarrow$ $\rightarrow$ $\mathcal{C}$ $\cong$ new.metab	boanalyst.ca/MetaboAnalyst/Secure/process/SanityCheck.xhtml	G 🖻 🖈 🖪 🕘 :
<b>#</b>	Data Integrity Check:	Show R. Commands
Upload ✓ Processing Data check Massing value Data filter Data differ Data differ Normalization Set parameter View result Download Exit	<ul> <li>1. electing the class labels - at lasts three reglicates are required in each class.</li> <li>1. if the samples are paired, the pair labels must conform to the segored contrast.</li> <li>1. if the samples are paired, the pair labels must conform to the segored samples of the samples are paired. The pair labels of fattures with constant values (a all areos).</li> <li>1. The presence of missing values or fattures with constant values (a all areos).</li> <li>1. Charge constant defaures in the samples of the samples of the samples of the samples are paired. The pair labels of the samples are paired with an expected pair labels of the samples of the samp</li></ul>	
		0

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 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).

seaboAnay	Metabo	<mark>Analyst 5.0</mark> - ข	iser-friendly, streamlined metabol	omics data analysis					
Contractions • Processing forta charge	Download Results & Please download the resu generate a PDF analysis r	Start New Journey ilts (tables and images) from the eport using the button. Finally, yo	Results Download tab below. The Download.zip contains all th u can continue to explore other compatible modules using the	ne files in your home directory. You can also Start New Journey tab.					
Set parameter	Results D	ownload Start New Journey							
View result Download	Gener	Generate Report							
Exit	Downloa	d.zip	data_original.csv						
	Rhistory,	3	mummichog_matched_compound_all.csv						
	data_pro	cessed.csv	scattermum.json						
	peaks_to	<u>paths 0_dpi72.png</u>	mummichog_pathway_enrichment.csv						
	mummic	hog_guery,json							
			Logout						

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).

new.metaboanalys	st.ca/MetaboAnalyst/ModuleView.xhtml					G	🖻 🛧 📙 🐐 🔲
Login to Canvas - U							
***	MetaboAnalyst !	.0 - user-friendly, stre	eamlined metabolomi	cs data analysis			
<u>^</u>	Module Overview						
	Input Data Type	Available Modules (click o	n a module to proceed, or scre	oll down for more details)			
<u>stR</u>	Raw Spectra (mzML, mzXML or mzData)			LC-MS Spec	tra Processing		
x	MS Peaks (peak list or intensity table)			Functional Analysis	Functional Meta-analysis		
	Annotated Features (compound list or table)		Enrichment Analysis	Pathway Analysis	Joint-Pathway Analysis	Network Analysis	
	Generic Format (.csv or .bxt table files)	Statistical Analysis [one factor]	Statistical Analysis [metadata table]	Biomarker Analysis	Statistical Meta-analysis	Power Analysis	Other Utilities
inada jébec	Statistical Analysis Ione face This module offers various common machine learning methods includ PLS-Da do drithogonal PLS-Da I clustering and valualization tools i and heatmaps as well as to classifi the statement of the statement of the statement and heatmaps as well as to classification.	eri hy used statistical and g t-tests, ANOVA, PCA, also provides ocreate dendrograms data based on andom	Statistical Anal This module aims to pherotypes and met considerations of ot based on general lin heatmaps for visual	ysis [mstadata table] detect associations between abolomics features with ere eperimental factors / covariates aer models coupled with PCA and zaton. More optioms are available for	>> This on r singl mett bior	Biomarker Analysis module performs various biomarker ar celevier operating characteristic ROCI ( e or multiple biomarkers using veller- och. It also allows users to manually sp arker models and perform new sample	halynes based urves for a tablished seefly
>	forests and SVM.		two-factors / time-se	rries data. sis (targeted)	>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>	Network Explorer	
ERC .							

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'.

∎ 5·∂· •										Excersis	e_4_multiomics.	csv - Excel									ور 🛦	mes McCul	lagh 🖭	-	a ×
File Home Insert Page Layout	Formulas D	ata Revie	w View	Help	ChemOff	ice21 I	Nuance PDF	Acrob	± Q1	iell me w	what you want to	do													S Share
Calibri	- 11 - A A	= = =	81.	🕫 Wrap Te	st	General		- 8		1	Normal	Bad		Good	Neu	tral	Calculat	ion		<b>&gt;</b>	∑ Auto	Sum - 🖌			
Paste Copy *					. C	- a		.00 Conc	itional Fo	mat as	Check Cell	Expla	aton	Innut	Link	ed Cell	Note		Insert	Delete Forma	Fill -	2 Se	ort & Find &		
V V Format Painter			E	- Merge o	s Center *	9 · 9	5 * 120	Form	atting * T	able *	cilieux cell	CAPIUI	intory	input		eu cen	Note	<b>T</b>	v	* *	🔹 🦑 Clea	r* Fi	Iter * Select *		
Clipboard rs Font	r 6		Alignme	ent	6	N	unber	ra					Styles							Cells		Editing			^
E1 * 1 × √ fe	log2(FC)																								^
Document Recovery	A A	8	C	D	E	F	G	н	1		J K	L	м	N	0	Р	Q	R	s	T	U	V	W	x	Y A
,	2 HMDB0000	7.9033	S	ATB2	10.936																				
Excel has recovered the following files. Save	3 HMDB0000	9.2548	н	LA-B	10.52																				
the ones you wish to keep.	4 HMDB0000	-3.6478	н	LA-C	10.319																				
	5 HMDB001	-1.7139	P	ECAM1	9.5302																				
TCA_JM_432_June2022.xlsx [	6 HMDB0000	-2.4953	S	CML1	9.3547																				
27/09/2022 14:09	7 HMDB000.	-1.4278	G	BP1	9.311																				
Untargeted Method Assess	8 HMDB000	0.85121	н	1-0	9.0044																				
Version created from the last	9 HMDB000	-0.961/4	C	INNA3	-8.8657																				
16/11/2022 16:48	11 HMDB000	-1.0938	0	D109	8 3615																				
Untargeted Method Assess	12 HMDB0000	-0 77727	Δ	RHGAPS	8 3597																				
X Version created last time the	13 HMDB0000	-0.84657	c	D4	8.2542																				
09/11/2022 16:26	14 HMDB000	0.57761	S	AMD5	-8.1552																				
KVL_1hr_3 conc_mero_13102	15 HMDB000;	-0.90158	н	IPK2	8.1349																				
X Version created from the last	16 HMDB0000	0.57974	U.	ACA	8.111																				
16/11/2022 16:48	17 HMDB0000	0.64706	N	ID1	7.685																				
KVL_1hr_3 conc_mero_13102	18 HMDB000	0.43207	IF	144	7.5658																				
A, Version created last time the	19 HMDB000:	1.2881	M	YO1F	7.3588																				
10/11/2022 12/45	20 HMDB0000	-1.3484	FL		7.351																				
Metabolomics Data Analysis	21 HMDB0000	0.65114	IG	F2BP3	6.4233																				
24/11/2022 11:45	22 HMDB000	-0.48804	P	SMB9	6.3854																				
	23 HMDB000:	0.41179	T	MSB4X	6.369																				
	24 HMDB006	0.82213	IT	GAM	6.1311																				
	25 HMDB0000	0.31323	P	LAGL1	6.0112																				
	26 HMDB000	0.35572	0	UHI	-5.9654																				
	28 HMDB000	0.58692	W	/AS	5 5966																				
	29 HMDB000	0.37711	A	LDH1A1	5.5618																				
	30 HMDB0000	0.41573	RI	HEX	-5.4809																				
	31 HMDB0000	0.29263	IF	ITM3	5.4012																				
	32 HMDB000:	0.7055	v	IM	5.1271																				
	33 HMDB0000	0.36242	PI	LXNC1	5.0946																				
	34 HMDB0000	0.21612	SE	ETD7	4.9714																				
	35 HMDB0000	-0.28271	SL	LCO2B1	4.9397																				
	36 HMDB0000	0.39337	T	HBS1	4.9331																				
	37 HMDB0000	0.38889	N	LRC5	4.7362																				
	38 HMDB000	0.2276	T/	AP1	4.7098																				
	39 HMDB0000	-0.18731	IF	ITM1	4.7035																				
	40 HMDB0000	0.31677	TS	SPAN13	-4.6816																				
	41 HMDB0000	0.32259	C	OCH	-4.6488																				
	42 HMDB0000	-0.33409	N	XF3	4.578																				
Which file do I want to save?	43 HMDB000	0.80322	0	TPN	-4.5635																				
Close	HWD5000.	Eventie	4 multion	nice	4.3360											: 4									
		EACCISIS	mundor		0											- LL						000 (m)			
Ready																						# (i)	H =		

#### Figure 22: Data found in Exercise\_4\_mulitomics.csv'

<ul> <li>↔ C</li> <li>A new.metaboanalyst</li> <li>⊗ New Tab</li> <li>♦ Login to Canvas - U</li> </ul>	st.ca/MetaboAnalyst/upload/JointUp	oloadView.xhtml		
Upload > Integrative Analysis Download Exit	Upload your data This module is designed for integra • Transcriptomics / Proteom • A list of significant ge • Metabolomis: • Targeted metabolomi • Untargeted metabolomi	itive analysis of transcriptomics i <b>tes:</b> anes (or proteins) with optional ics: a list of significant metabol omics: a <u>complete</u> peak list. Pea	/proteomics and metabolomics data fold change values; ites (or compounds of interest) with k names must be their numeric mass	a <u>t pathway level</u> . optional fold change values; s (m/z) values with optional retention times;
	Organism: Homo sapiens	; (human) 🗸	Metabolomics Type:	Targeted (compound list) V
	Genes/proteins with optic	onal fold changes	Compound list with o	ptional fold changes
	SATD2 10.936 HLA-8 10.52 HLA-C 10.319 PECAM1 9.3302 SCML1 9.3547 GBP1 9.311 H1-0 9.0044 CTINNA3 -8.8657 RHOU 8.4613 CD109 8.3615 ARHGAP5 8.3597 CD4 8.2542 SAMD5 -8.1552 HIPK2 8.1349 UACA 8.111 ND1 7.685	, Symbol 🗸	HMDB0000710           HMDB000025           HMDB0000225           HMDB0001220           HMDB0001259           HMDB0001259           HMDB0001259           HMDB0001259           HMDB0001259           HMDB0001259           HMDB0000167           HMDB0000539           HMDB0000539           HMDB0000539           HMDB0000531           HMDB000230           ID Type:           HMDB ID	7.9033 9.2548 3.6478 -1.7139 -2.4953 -1.4278 0.87121 -0.96174 -0.87956 -1.0938 -0.57761 -0.90158 0.577761 -0.90158 0.57774 0.64706
			Submit Try our	example

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'

	The system requires all the IDs approximate match by clicking	s (except common co g the <b>View</b> link in the	ompound names) to be ma e Details column. To <b>remov</b>	tched exactly. The table below shows the matched genes and compounds from the underlying databases. For common compound e a gene or compound from further analysis, use the <b>Delete</b> link in the last column.	names, users can further perform
Upload	Gene Name Mapping	Compound Nam	ne Mapping		
ID map					
Set parameter View result	Query	Hit	Symbol	Name	
Download	SATB2	23314	SATB2	SAT8 homeobox 2	Delete
Exit	HLA-B	3106	HLA-B	"major histocompatibility complex, class I, B"	Delete
	HLA-C	3107	HLA-C	"major histocompatibility complex. class I. C"	Delete
	PECAM1	5175	PECAM1	platelet and endothelial cell adhesion molecule 1	Delete
	SCML1	6322	SCML1	Scm.polycomb.group.protein.like_1	Delete
	GBP1	2633	GBP1	guanylate binding.protein 1	Delete
	H1-0	3005	<u>H1-0</u>	H1.0 linker histone	Delete
	<b>CTNNA3</b>	29119	CTNNA3	catenin alpha 3	Delete
	RHOU	58480	RHOU	ras homolog family member U	Delete
	CD109	135228	CD109	CD109 molecule	Delete
	ARHGAP5	394	ARHGAP5	Rho GTPase activating protein 5	Delete
	CD4	920	CD4	CD4 molecule	Delete
	SAMD5	389432	SAMD5	sterile alpha motif domain containing 5	Delete
	HIPK2	28996	HIPK2	homeodomain interacting protein kinase 2	Delete
	UACA	55075	UACA	uveal autoantigen with coiled-coil domains and ankyrin repeats	Delete
	NID1	4811	NID1	nidogen 1	Delete
	1F144	10561	<u>IEI44</u>	interferon induced protein 44	Delete
	MYO1F	4542	MYO1F	myosin IE	Delete
	FLI1	2313	<u>FU1</u>	"Fli-1 proto-oncogene, ETS transcription factor"	Delete

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.

<b>A</b>	Click the corresponding Pathway Name to view to graphical presentation, click Match Status to view the pathway members (with matched ones highlighted).													
Upload	بل Results Table	🛃 Matched Fe	atures		▶ Net	work Explorer								
<ul> <li>Integrative Analysis</li> <li>ID map</li> </ul>	Pathway Name	Match Status	p value	-log(p)	Holm p	FDR	Impact	Link						
Set parameter	Citrate cycle (TCA cycle)	21/42	1.5623E-9	8.8062	1.3123E-7	1.3123E-7	2.2195	KEGG						
View result	Pyruvate metabolism	19/45	3.0129E-7	6.521	2.5007E-5	1.2654E+5	1.4545	KEGG						
Download	Pentose phosphate pathway	19/47	6.8377E-7	6.1651	5.6069E-5	1.9145E-5	2.0652	KEGG						
Exit	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	0.005635	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						
		~~ ~	1 2 3 4	> >>										
			Proceed											

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?