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

- 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).

ormats							
	Input Data Type	Available Modules (click on a	module to proceed, or scroll down	for more details)			
da	Raw Spectra (mzML, mz/OML or mzData)			LC-MS Spect	ral Processing		
History	MS Peaks (peak list or intensity table)			Functional Analysis	Functional Meta-analysis		
<u>aAnalystR</u>	Annotated Features (compound list or table)		Enrichment Analysis	Pathway Analysis	Joint-Pathway Analysis	Network Analysis	
<u>tats</u>	Generic Format (.csv or .td table files)	Statistical Analysis	Biomarker Analysis	Time-server	Statistical Meta-analysis	Power Analysis	Other Utilities
tions				Show R command	history		
					,		
omeCanada	O Statistical Analysis		Biomarker Analysis			way Analysis (targated)	
ome Canada	This module offers various commonly used s		This module performs vario	us biomarker analyses based on	© Past This m	odule supports pathway analysis (integrating e	
	This module offers various commonly used s machine learning methods including I-tests, A DA and Orthogonal PLS-DA, it also provides	NOVA, PCA, PLS- clustering and	This module performs vario receiver operating characte multiple biomarkers using w	eristic (RDC) curves for a single or reli-established methods. It also	D put This m analys 26 mc	odule supports pathway analysis (integrating e is and pathway topology analysis) and visualiz sel organisms, including Human, Mouse, Rat, Co	zation for ow,
	This module offers various commonly used s machine learning methods including t-tests, A	NOVA, PCA, PLS- clustering and d heatmaps as well	This module performs vario receiver operating characte multiple biomarkers using w	eristic (ROC) curves for a single or vell-established methods. It also secify biomarker models and	D Pat This m analyz 25 mo Chicks	odule supports pathway analysis (integrating e is and pathway topology analysis) and visualiz	zation for ow,
	This module offers various commonly used is machine learning methods including 1-tests, A DA and Orthogonal (FS-DA), it also provides visualization tools to create dendrograms an	NOVA, PCA, PLS- clustering and d heatmaps as well	This module performs vario receiver operating characte multiple biomarkers using w allows users to manually sp	eristic (ROC) curves for a single or ell-established methods. It also beeily biomarker models and tion.	C but The m analysy 28 may Choice Matern	odule supports pathway analysis (integrating e is and pathway topology analysis) and visualz sel organisms, including Human, Mouse, Rat, Cc n, Zebra fish, Arabidopsis thaliana, Rice, Dros	zation for ow,
nomeCanada nomeQuébec	The module offers service commonly used is machine teaming methods including Lesss, A DA and Omogoni R2-DA. Kate provides voluciation too is to create development an as to creatly data based on rankom forests O Sector Advantasi This module above savers to uplaad raw (CA	NGVA, PCA, PCS. Cutatoring and d SVII. AS aspectra (ms)/I.,	This module performs vario receiver operating characterist mutiple biomativer using via perform new sample predor O Euclidonal Analysis, IMS This module accepte high-re-	Instic (ROC) curves for a single or el-stabilished methods. It also seefly biomarker models and tool. Phatas) resolution LC-MS spectral peak data	 ○ Fair This in analy 28 may 28 may 29 may 20 may 20	odule supports path way analysis (integrating e and pathway hopology analysis) and visualize del organismi, including human, Mouse, Rai, CC, m., Zebrafane, including human, Mouse, Rai, CC, S. & cerevise, <i>E. Coll,</i> and others species. 	zation for ovy, ophila,
iomeQuébec	This module offers various commonly used a machine learning methods including takets, a Da and dribogonal R-S-DA. Take provides visualization look to create development an as to classify data based on random forests C. Scectral Analysis	NOVA, PGA, PLS. Quatering and Indentings as well and SVM. AS spectra (mzMu, r optimized	This module performs varior receiver operating characti- mutiple bomatries using va- allevs users to manually users to manually perform new sample predic Definitional Analysis //155 This module accepts high-r to perform metabolic pable	ristic (ROC) curves for a single or ell-established methods. It also secify biomarker models and toon.	© the analysis 20 the Other Waters Then the	odule supports pathway analysis (integrating es and pathway popology analysis) and visualiz etel organisms, including Human, Mouse, RLC 2. Zebartshi, Androhoginz itahiana, Rich, Dros I, S. cerevisae, E. Coll, and others species.	zation for ow, ophita, onsistent mics
	The module offers various commonly used machine saming embody activity of DLA and offorgan FT-S-D-L table privates visualization tobits to based on random formats of statistical characteristic for statistical characteristic The module activity assets to splited ensure to mobility or mobiles to be proceeded using an embolity of the module activity and the splited ensure to the mobility of the module activity and the splited ensure to mobility or mobility to be proceeded using an embolity of the splited ensure to activity of	MOUAL PCR PLG- clustering and beforeps as well and SVM. AS spectra (mdRL, rightmad XIS The moduli pack Intensity	This module performs varior receiver operating characti- nutgible biometines varies was perform metive sample predi- or <u>functional Analysis</u> . This module accepte high-ri- to perform metalosic pathwi- exploration based on the vari-	Institle (ROC) curves for a single or e8-astabilished methods. It also obly blomarier models and too: Pastability esolution (C-MS spectral peak data ay encliment analysis and vasal e4-attability for a discrimination fra discrimination for the discrimination fra discrimination for the discrimination for the discrimination of the discrimination for the discrimination for the discrimination of the discrimination for the discrimination of the discrimination for the discrimination of the discrimination	© Par This m analy 20 may Check Marris This m function	odule supports pathways analysis (integrating et and pathway hotpology analysis) and visualize del organisma, incluiding Human, Noise, Rat, CC, J. Zehrefah, Anabiotes Inhian, Rice, Dre- a, S. cerevisae, <i>E. coli</i> , and others species. citized Meta-analysis (115 ceato) dula provides statistical methode to identify co and changes correct smalle pobla methodor.	cation for ow, ophile, unsistent mics ss. t. ent s 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
Lipicad Processing	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 - cachesic; 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 Data Format: Samples in columns Data File: Browse Exercise_2_DATA_MA.csv Use the example data Data Description Datasel Urinary metabolite concentrations from 77 cancer patients measured by 1H NIMR. Phenotype: N-cachexic; Y - control 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).

ñ	Data Integrity Check:
	 Checking the class labels - at least three replicates are required in each class.
Upload	If the samples are paired, the pair labels must conform to the specified format.
Processing	3. The data (except class labels) must not contain non-numeric values.
Data check	The presence of missing values or features with constant values (i.e. all zeros).
Name check	
Missing value	Data processing information:
Data filter Data editor	Checking data contentpassed.
Normalization	Samples are in columns and features in rows.
Download	The uploaded file is in comma separated values (.csv) format.
Exit	The uploaded data file contains 18 (samples) by 109 (compounds) data matrix.
LAIL	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, missing values will be replaced by 1/5 of min positive values of their corresponding variables
	Click the Skip button if you accept the default practice;
	Or click the Missing 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	10542	<u>C00492</u>	
ne check sing value	HMDB0001548	D-Ribose 5-phosphate	HMDB0001548	439167	<u>C03736</u>	
a filter	HMDB0000618	D-Ribulose 5-phosphate	HMDB0000618	439184	<u>C00199</u>	
a editor	HMDB0000792	Sebacic acid	HMDB0000792	<u>5192</u>	<u>C08277</u>	
alization	HMDB0060274	Sedoheptulose 1,7-bisphosphate	HMDB0060274	<u>164735</u>	<u>C00447</u>	
oad	HMDB0060509		-	-	-	View
	HMDB0001068	D-Sedoheptulose 7-phosphate	HMDB0001068	22833559	<u>C05382</u>	
	HMDB0000247	Sorbitol	HMDB0000247	5780	<u>C00794</u>	
	HMDB0005831	Sorbitol-6-phosphate	HMDB0005831	618	<u>C02810</u>	
	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	<u>17473</u>	C00167	
	HMDB0000286	Uridine diphosphate glucose	HMDB0000286	<u>53477679</u>	<u>C00029</u>	
	HMDB0000300	Uracil	HMDB0000300	<u>1174</u>	C00106	
	HMDB0000289	Uric acid	HMDB0000289	1175	<u>C00366</u>	
	HMDB0000296	Uridine	HMDB0000296	6029	<u>C00299</u>	
	HMDB0000288	Uridine 5'-monophosphate	HMDB0000288	6030	C00105	
	HMDB0000290	Uridine diphosphate-N-acetylglucosamine	HMDB0000290	9547196	C00043	
	HMDB0000285	Uridine triphosphate	HMDB0000285	6133	<u>C00075</u>	
	HMDB0001554	Xanthylic acid	HMDB0001554	73323	C00655	
	HMDB0002917	D-Xyiitol	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	
	HMDB0000125			-	-	View

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.
lame check Missing value Data filter Data editor	Sample Normalization
malization vnload	Sample-specific normalization (i.e. weight, volume) Specify I formalization by sum
	Normalization by median Normalization by reference sample (PQN) Specify
	Itermalization by a pooled sample from group Specify Itermalization by reference feature Specify
	Quantile normalization Data transformation
	Log transformation (generalized logarithm transformation or glog) Cube root transformation (takes 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) Pareto 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'

staboAnays 5.0	MetaboAnalys	t 5.0 - user-friendly, streamlined metabolon	nics data analysis
ñ	Specify pathway analysis parameter	'S:	
Upload Processing	Visualization method	Scatter plot (testing significant features) Heatmaps (testing your selected features)	
a check ne check sing value	Enrichment method	 Global Test Global Ancova 	
le la	Topology analysis	Relative-betweeness Centrality Out-degree Centrality	
	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)	
	Mammals	Homo sapiens (KEGG) Homo sapiens (SMPDB) Mus musculus (KEGG) Mus musculus (SMPDB) Rattus norvepicus (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) Ilasmodium falciparum 307 (Malaria) (KEGG) Plasmodium vivax (Malaria) (KEGG) Trypanosoma brucel (KEGG)	
		Escherichia coli K-12 MG1655 (KEGG) Bacilus subtilis (KEGG) Pseudomonas putida KT2440 (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

â	Result View:								
Upload	The metabolome view on the left shows all matched pathways according to the p values from the pathway enrichment analysis and pathway impact values from the pathway topology analysis. Placing your mouse over each pathway node will reveal its pathway name. <u>Clicking each node</u> will launch the pathway view on the right panel								
Processing Data check Name check	The pathway can be launched either by clicking the corresponding node on the left image or by clicking the pathway name from the table below. Please note, each node (compound) is clickable. You can <u>zoom in and out</u> using the control buttons below, and then <u>drag</u> the image to the locations of interest. Placing the <u>mouse over</u> each metabolite node will reveal its common name. <u>Clicking the node</u> will trigger the compound view of the selected compound.								
Missing value Data filter Data editor	metabolite is not in you	data and is also excluded fr		y applicable if you have uplo		chment analysis; <u>grey</u> means profile); other colors (varying fi			
Normalization Download	Show gridline	Update	26			26 12			
Exit		Overview of Pathway Ana	lysis	Ubiquinor	ne and other terpenoid-quin	one biosynthesis			
				C00082	C00828	C03313 C05849 C02059			

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).

	1-4	<4 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	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
	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.

1 e ee 1 2 DO DI										
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	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	<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	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			
<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	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 10: Download and save report for pathways analysis.

solaboAna/	N	/letaboAnalyst	5.0 - user-fi	riendly, streamlined metabo	lomics data analysis
Upload • Processing • Massing value Data filter Data filter Data filter Data filter Bata	Please dov	PDF analysis report using the butto	es) from the Results D	Download tab below. The Download.zip contains all tinue to explore other compatible modules using the modules using the eath.view_0_dei72.eng data_normalized.csy Lysine degradation.eng data_original.csy norm_0_dei72.eng snorm_0_dei72.eng pathway_results.csy	

Figure 11: Generate a report

10

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	\$	*	• •	:
	Dynosing Staparanter View result Download Exit		This r <u>as de</u> •	fined by metabolite sets or The input peak list or peak [Required] Feature or peak [Optional] You can also pro	pathways This is because changes at grou. table must contain the <u>comistic data</u> , not manes must be their <u>munnets must more</u>), vixide retention time (RT) to further improve sak intensity table	p level rely on "collective behavior" which just significant data - we need the comple <u>values</u> for putative annotation;	is more tolerant to random ete data to estimate the null	essumption is that <u>outative annotation at individual</u> errors in compound annotation as demonstrated by model (background):	at function	levels			
				Try our example dat. Data	Subr asets Format	nt Description							
				O Immune Ce	retention time)	Example peak intensity table from KO dendritic cells and epithelial cells treat Peak intensity table of COVID-19 glob.	ted in DSS.						
				Covid-19	Peak intensity table with retention 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	🗴 🗇 Peak Annotation and Verification 🗴 G. branched chain amino acids - Go 🗴 🔹 MetaboAnalyst 🗙 🕇	∨ - ∂ ×
\leftrightarrow \rightarrow \mathfrak{C} $\ $ new.metab	zoanalyst.ca/MetaboAnalyst/Secure/process/SanityCheck.xhtml	G 🖻 🖈 🖬 🕘 🗄
#	Data Integrity Check:	Show R. Commands
Upload ✓ Processing Data direck Missing value Data filter Data silter Data silter Data silter Data silter Normalization Se parameter View result Download Exit	 1. election the class labels - at least three regulates are required in each class. 2. If the samples are pared, the pair labels must contram to the specified format. 3. If the samples are pared, the pair labels must contram to unless (a. all acros). 3. The presence of missing values or features with constant values (a. all acros). Checking data content - pared. 3. Bayes are in columns and features in root. 3. The uploaded data in a contain a values (a. all acros). 3. The uploaded data in a contain and values (a. all acros). 3. The uploaded data in a contain and values (a. all acros). 3. Suppose are discided in a contain samples by 0.724 (pesk)ma;//tj) data matrix. 3. Suppose were discided in a contain samples by 0.724 (pesk)ma;/tj) data matrix. 3. Suppose were discided in the contain sit (amples) by 0.724 (pesk)ma;/tj) data matrix. 3. Suppose were discided in a contain sit (amples) by 0.724 (pesk)ma;/tj) data matrix. 3. Suppose were discided in the contain sit (amples) by 0.724 (pesk)ma;/tj) data matrix. 3. Suppose were discided in the contain sit (amples) by 0.724 (pesk)ma;/tj) data matrix. 3. Suppose were discided in the contain sit (amples) by 0.724 (pesk)ma;/tj) data matrix. 3. Suppose were discided in the contain site (angles across samples were found and distet. 4. Stati data 0.0101 maining values were discided in the corresponding matrix) is a discided in the found in the subscided of the distet of a matrix. 3. Class are supposed by 0.124 mining values were discided in the supposed of the corresponding matrix) is a discided in the supposed of the corresponding matrix is a discided in the supposed by 0.124 mining values were discided in the supposed by 0.124 mining values were discided in the supposed by 0.124 mining values in the supposed by 0.124 mining values is a discided by 0.124 mining values in the supposed by 0.124 mining values	
		•
	Xia Lab @ McQ8 (Instructioned 2022-11-17)	U

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.

 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).

seaboAnalys	Meta	boAnalyst <mark>5.0</mark> -	- user-friendly, streamlined metabol	omics data analysis
Cupical - Processing Data check	Please download the		the Results Download tab below. The Download.zip contains all t , you can continue to explore other compatible modules using the	
Set parameter	Resu	Its Download Start New Journ	ey	
View result Download	Ge	enerate Report		
Exit	Dow	nload.zip	data_original.csv	
	Rhis	tory.R	mummichog_matched_compound_all.csv	
	data	processed.csv	<u>scattermum.json</u>	
	peak	<u>s to paths 0 dpi72.png</u>	mummichog_pathway_enrichment.csv	
	mun	michog_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).

gin to Canvas - U	MetaboAnalyst/ModuleView.xhtml					G	ie 🛧 📕 🗯 🔲
	MetaboAnalyst	5.0 - user-friendly, stre	eamlined metabolomi	cs data analysis			
м	odule Overview						
	Input Data Type	Available Modules (click o	on a module to proceed, or scro	oll down for more details)			
B	Raw Spectra (mzML, mzXML or mzData)			LC-MS Spec	tra Processing		
	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 .txt table files)	Statistical Analysis [one factor]	Statistical Analysis [metadata table]	Biomarker Analysis	Statistical Meta-analysis	Power Analysis	Other Utilities
ada bec	Statistical Analysis (one fact This module offers various comm machine learning methods include PLS-DA and Orthogonal PLS-DA. I clustering and vasualization tools and heatmaps as well as to classifi forests and SVM.	only used statistical and ng t-tests, ANOVA, PCA, t also provides o create dendrograms	This module aims to phenotypes and met considerations of ot based on general lin	ysis (metadata table) defect associations between abdomics features with ther experimental factors / covariates are models coupled with PCA and ation. More options are available for rise data.	This n on ree single meth	Ilonarker Analysis nodule performs various biomarker a even operating characteristic (ROC) or multiple biomarkers using well- dist. It also allows users to manually s raker models and perform new sampl	curves for a stablished specify
RC VG	>> Enrichment Analysis		>> Pathway Analy	sis (targeted)	»» <u>F</u>	letwork Explorer	

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'.

🗄 5-0-+								Excersi		sv - Excel									🐴 Jami	es McCulli	igh 🖭		
ile Home Insert Page Layout	Formulas Dat	a Review	View Help	ChemOffice	21 Nuance	PDF Ac	robat 🤇	7 Tell me	what you want to e	do													R₄ Shar
Calibri •	11 * A* A*		♡ - 한 Wrap Ti	ext .	General		,		Normal	Bad	C	Good	Neutr	ral	Calculati	ion	•	>	∑ AutoSu	um * A	v		
ste Servat Painter	- <u>A</u> -	= = = =	🗉 💓 Merge	& Center 👻	ş. % ,	18 48 C		Format as	Check Cell	Explan	tory	nput	Unke	d Cell	Note	*	Insert	Delete Format	🐺 Fill - 🖉 Clear -	So	t& Find & er∗ Select∘		
Clipboard 12 Font	6		Alignment	6	Number		irmatung *	HOIG +			Styles						· ·	Cells		Editing	er • Select •		
	g2(FC)																						
ocument Recovery	A A Metabolit lo		C D Transcript	E	F G	н	1		J K	L	м	N	0	P	Q	R	S	т	U	V	W	х	Y
ocument necessary	2 HMDB000		SATB2	10.936																			
cel has recovered the following files. Save	3 HMDB000	9.2548	HLA-B	10.936																			
e ones you wish to keep.		9.2548	HLA-B HLA-C	10.52																			
TCA_JM_432_June2022.xlsx [5 HMDB001		PECAM1	9.5302																			
Version created last time the		-2.4953	SCML1	9.3547																			
27/09/2022 14:09		-1.4278	GBP1	9.311																			
Untargeted Method Assess	8 HMDB0000		H1-0	9.0044																			
	9 HMDB000		CTNNA3	-8.8657																			
Version created from the last 16/11/2022 16:48	10 HMDB000:		RHOU	8.4613																			
	11 HMDB000.		CD109	8.3615																			
Untargeted Method Assess	12 HMDB0000		ARHGAP5	8.3597																			
Version created last time the 09/11/2022 16:26	13 HMDB0000	-0.84657	CD4	8.2542																			
	14 HMDB000	0.57761	SAMD5	-8.1552																			
KVL_1hr_3 conc_mero_13102	15 HMDB000:	-0.90158	HIPK2	8.1349																			
Version created from the last	16 HMDB0000	0.57974	UACA	8.111																			
16/11/2022 16:48	17 HMDB0000		NID1	7.685																			
KVL_1hr_3 conc_mero_13102	18 HMDB000		IE144	7.5658																			
a. Version created last time the	19 HMDB000:		MY01F	7.3588																			
a, 16/11/2022 12:43	20 HMDB000		FUI	7.351																			
Metabolomics Data Analysis	20 HMDB0000		IGF2BP3	6,4233																			
•••	22 HMDB000:		PSMB9	6.3854																			
	23 HMDB000:		TMSB4X	6.369																			
	24 HMDB0060	0.82213	ITGAM	6.1311																			
	25 HMDB0000	0.31323	PLAGL1	6.0112																			
	26 HMDB000	0.35572	CDH1	-5.9654																			
	27 HMDB0000	0.38062	IFI16	5.634																			
	28 HMDB000	0.58692	WAS	5.5966																			
	29 HMDB0000	0.37711	ALDH1A1	5.5618																			
	30 HMDB0000		RHEX	-5.4809																			
	31 HMDB0000		IFITM3	5.4012																			
	32 HMDB000:	0.7055	VIM	5.1271																			
	33 HMDB000		PLXNC1	5.0946																			
	34 HMDB0000		SETD7	4.9714																			
	35 HMDB0000		SLCO2B1	4.9397																			
	36 HMDB000		THBS1	4.9331																			
	37 HMDB0000		NLRC5	4.7362																			
	38 HMDB0000	0.2276	TAP1	4.7098																			
	39 HMDB0000		IFITM1	4.7035																			
	40 HMDB0000		TSPAN13	-4.6816																			
	41 HMDB0000	0.32259	COCH	-4.6488																			
	42 HMDB0000	0.33409	NXF3	4.578																			
	43 HMDB000		DNAJA4	-4,5635																			
Which file do I want to saver		0.27753	UTRN	4.5586																			
Close	< >	Excersise_4		+																			

Figure 22: Data found in Exercise_4_mulitomics.csv'

roteins) with optiona of significant metabo	s/proteomics and metabolomics data a I fold change values; lites (or compounds of interest) with o ak names must be their numeric mass i	ptional fold change values;	ntion times;
changes	Metabolomics Type: Compound list with op	Targeted (compound list) V	
-	НМDB0000694 НМDB0000694 НМDB0000225 НМDB0000228 НМDB0000208 НМDB0000208 НМDB0000399 НМDB0001548 НМDB0001548 НМDB0001547 НМDB0000355 НМDB000025831 НМDB0000247 НМDB0000247 НМDB0000247 НМDB0000247	7.9033 9.2548 -3.64778 -1.7139 -2.49533 -1.4278 0.85121 -0.96174 -0.87956 -1.70938 -0.77727 -0.84657 0.57761 -0.90158 0.57974 0.64706	
	* //	ID Type: HMDB ID	HMDB0000230 0.64706

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'

				tched exactly. The table below shows the matched genes and compounds from the underlying databases. For commo e a gene or compound from further analysis, use the Delete link in the last column.	n compound names, users can further perfo
Upload Integrative Analysis	Gene Name Mapping	Compound Name	e Mapping		
ID map Set parameter View result	Query	Hit	Symbol	Name	
Download	SATB2	23314	SATB2	SATB homeobox 2	Delete
Exit	HLA-B	3106	HLA-B	"major histocompatibility complex, class J. B"	Delete
	HLA-C	3107	HLA-C	"major histocompatibility complex_class 1. C"	Delete
	PECAM1	<u>5175</u>	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
	CTNNA3	29119	CTNNAS	<u>catenin alpha 3</u>	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>IFI44</u>	interferon induced protein 44	Delete
	MYO1F	4542	MYO1E	myosin IE	Delete
	FLI1	2313	FLI1	"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.

enter uit Pentose Glycolys Lysine d Pyrimidi Glytabis Inssite[Nitroger Glycolys	y Name del (TCA cycle) metabolism phosphate pathway s or Gluconeogenesis s or Gluconeogenesis gadatan ne metabolism ne metabolism metabolism metabolism	Match Status 21/42 19/43 19/47 22/61 16/49 24/99 15/56 15/56 5/10	p value 1.5623E-9 3.0129E-7 6.8377E-7 9.172E-7 1.2044E-4 5.0757E-4 0.0002298 0.00027876	-log(p) 8.8062 6.521 6.1651 6.0375 3.9192 3.2945 2.6925 2.5548	Holm p 1.3123E-7 2.5007E-5 5.4069E-5 7.4283E-5 0.0006349 0.040098 0.15832 0.21465	FDR 1.3123E-7 1.2654E-5 1.9145E-5 1.9261E-5 0.0020233 0.007106 0.024358	Impact 2.2195 1.4545 2.0652 1.3833 0.4375 1.398 0.61818
Uit Pyruvate Pyruvate Glycolys Lysine d Pyrimidi Glutathis Inositel Nitroger Glycoyta	metabolism phosphate pathway s or Gluconeogenesis garadation e metabolism phosphate metabolism metabolism	10/45 10/47 22/61 16/49 24/99 13/56 17/69	3.0129E-7 6.8377E-7 9.172E-7 1.2044E-4 5.0757E-4 0.0020298 0.0027876	6.521 6.1651 6.0375 3.9192 3.2945 2.6925	2.5007E-5 5.6069E-5 7.4293E-5 0.0096349 0.040098 0.15832	1.2654E-5 1.9145E-5 1.9261E-5 0.0020233 0.007106 0.024358	1.4545 2.0652 1.3833 0.4375 1.398
Pentose Glycolys Lysine d Pyrimidi Glutathi Inositol Nitroger Glycoyla	shosphate pathway s or Gluconeogenesis gradation e metabolism hosphate metabolism metabolism	19/47 22/61 16/49 24/99 15/56 17/69	6.8377E-7 9.172E-7 1.2044E-4 5.0757E-4 0.0020298 0.0027876	6.1651 6.0375 3.9192 3.2945 2.6925	5.6069E-5 7.4293E-5 0.0096349 0.040098 0.15832	1.9145E-5 1.9261E-5 0.0020233 0.007106 0.024358	2.0652 1.3833 0.4375 1.398
Glycolys Lysine d Pyrimid Glutathi Inositel Nitroger Glycoyla	s or Gluconeogenesis gradation ne metabolism ne metabolism metabolism metabolism	22/61 16/49 24/99 15/56 17/69	9.172E-7 1.2044E-4 5.0757E-4 0.0020298 0.0027876	6.0375 3.9192 3.2945 2.6925	7.4293E-5 0.0096349 0.040098 0.15832	1.9261E-5 0.0020233 0.007106 0.024358	1.3833 0.4375 1.398
Lysine d Pyrimidi Giutathi Inosito I Nitroger Głyczyła	gradaton e metabolism ne metabolism hosphate metabolism metabolism	16/49 24/99 15/56 17/69	1.2044E-4 5.0757E-4 0.0020298 0.0027876	3.9192 3.2945 2.6925	0.0096349 0.040098 0.15832	0.0020233 0.007106 0.024358	0.4375
Pyrimidi Glutathi Inositol Nitroger Głyoxyla	metabolism ne metabolism zhosphate metabolism metabolism	24/99 15/56 17/69	5.0757E-4 0.0020298 0.0027876	3.2945 2.6925	0.040098	0.007106	1.398
Glutathi Inositol Nitroger Glyoxyla	ine metabolism vhosphate metabolism metabolism	15/56	0.0020298	2.6925	0.15832	0.024358	
Inositol Nitroger Głyoxyla	hosphate metabolism metabolism	17/69	0.0027876				0.61818
Nitroger Glyoxyla	metabolism			2.5548	0.21465		
Glyoxyla		5/10				0.02927	0.48529
	e and dicarboxylate metabolism		0.0038734	2.4119	0.29438	0.036152	0.88889
Glycerol		14/56	0.005635	2.2491	0.42262	0.047334	0.89091
	pid metabolism	10/35	0.0068886	2.1619	0.50976	0.052604	0.64705
Pentose	and glucuronate interconversions	9/32	0.011393	1.9434	0.83167	0.078261	0.87097
Arginine	biosynthesis	8/27	0.012112	1.9168	0.87205	0.078261	0.53846
Aminoa	yl-tRNA biosynthesis	16/74	0.013874	1.8578	0.98508	0.083246	0.41096
Cysteine	and methionine metabolism	15/71	0.020794	1.6821	1.0	0.11645	0.61429
Terpeno	d backbone biosynthesis	9/36	0.02469	1.6075	1.0	0.12469	0.88571
Amino s	igar and nucleotide sugar metabolism	16/79	0.025234	1.598	1.0	0.12469	0.69231
Alanine,	aspartate and glutamate metabolism	13/61	0.028404	1.5466	1.0	0.13255	0.58333
Fructose	and mannose metabolism	9/40	0.046448	1.333	1.0	0.20535	0.74359
Butanoa	e metabolism	7/29	0.053966	1.2679	1.0	0.22666	0.60714

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

22

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?