Advanced Mass Spectrometry 1KB159
Sireethorn Poomborplab

XCMS Online and MetaboAnalyst

1. Introduction

Global metabolomics or untargeted metabolomics analysis is a widespread method of investigating the metabolic profile and pathway in various health-related conditions. Mass Spectrometry is a powerful analytical technique for metabolomic analysis. However, because of the complexity of the data, data treatment and interpretation between MS data and functional insight can be challenging and lead to misinterpretation if it is not done correctly.

A typical MS global metabolomics data analysis workflow starts by processing raw spectra to obtain each feature's mass spectra and retention time. Then, the metabolite is identified using various statistical tools and databases. This is followed by pathway analysis to integrate experiment data into a biological context.[1]

In this work, LC-MS metabolomics data were analysed using XCMS online tools, and statistical analysis was performed. This experiment aims to identify the altered endogenous metabolite and its biological pathway, which corresponds to gene knockdown of a mouse cell line compared to the wild type.

2. Metabolomics analysis using XC-MS online

2.1 Annotation of endogenous metabolite

The LC-MS analysis was performed in gene-knockdown (KO), wild type (WT) and Quality control (QC) groups of mice cell lines using Q-TOF as a mass analyser. The LC-MS data, including retention time, intensity and mass spectrum, was analysed with the XCMSOnline platform to compare the change in metabolite level between gene-knockdown (KO) and wild-type (WT) groups. The volcano plot was done to confirm the metabolite change, and the data was auto-scaled (mean-centred and divided by the standard deviation of each variable) for better interpretation. The compound that matches the criteria *(see Table 1 for criteria)* was annotated for further pathway analysis.

Another statistical analysis, including PCA and heat map analysis, was performed to confirm group differences and quality control checks.

Table 1 *Inclusion and Exclusion criteria for compound identification*

Inclusion criteria	Value	Exclusion criteria	Value
Mass error	≤ 3 ppm	Database information	- Exogeneous compound
Fold change	≥ 2		- Not present in HMDB or KEGG database
TIC	Visually different in chromatogram		
p-value	≤ 0.001 (in XCMS)		
	≤ 0.01 (in volcano plot)		

2.2 Pathway analysis

Pathway analysis was performed in all annotated compounds using MetaboAnalyst 6.0 to integrate the experiment data into the biological process. The result was correct with False Discovery Rate (FDR) and visualised by comparing it with the *Mus musculus* (house mouse) KEGG Library.

3. Top 5 altered endogenous metabolites

From the LC-MS data from XCMS and the volcano plot (Figure 2) that matched the criteria, 37 compounds were annotated and sorted by statistical significance (Table 2). The top 5 altered endogenous metabolites from highest to lowest significance level were 3-Methoxytyrosine, 2-(a-Hydroxyethyl)thiamine diphosphate, (1aalpha,2beta,3alpha,11calpha)-1a,2,3,11c-Tetrahydro6,11dimethylbenzo[6,7]phenanthrol [3,4-b] oxirene-2,3-diol, S-Lactoylglutathione and Leucyl-leucine.

4. Summary of pathway analysis

The result of pathway analysis reveals that there are 12 metabolism pathways involved in this experiment. The three most significant pathways are Riboflavin metabolism, Pyruvate metabolism and glutathione metabolism, as shown in Table 3 and Figure 4.

The downregulation of Riboflavin (Vitamin B2) and flavin mononucleotide were observed in the riboflavin metabolism pathway (Figure 5). In the Pyruvate metabolism pathway, the downregulation of 2-(alpha-Hydroxyethyl)thiamine diphosphate and the upregulation of S-D-Lactoylglutathione (Figure 7).

From overall observation, the change might be caused by a change in the starting compound biosynthesis, or the enzyme expression change in the metabolism pathway. Due to the different gene expressions, this can also be a biological response to stress, such as inflammation or oxidative stress. To increase the reliability of the result, it should be confirmed with another method, such as a spiking experiment, compared with another library or MSn analysis.

5. Reference

1. Lu, Y., Pang, Z. & Xia, J. Comprehensive investigation of pathway enrichment methods for functional interpretation of LC–MS global metabolomics data. *Brief. Bioinform.* **24**, bbac553 (2023).

6. Supplement chapter

6.1 Statistical analysis

6.1.1 PCA analysis

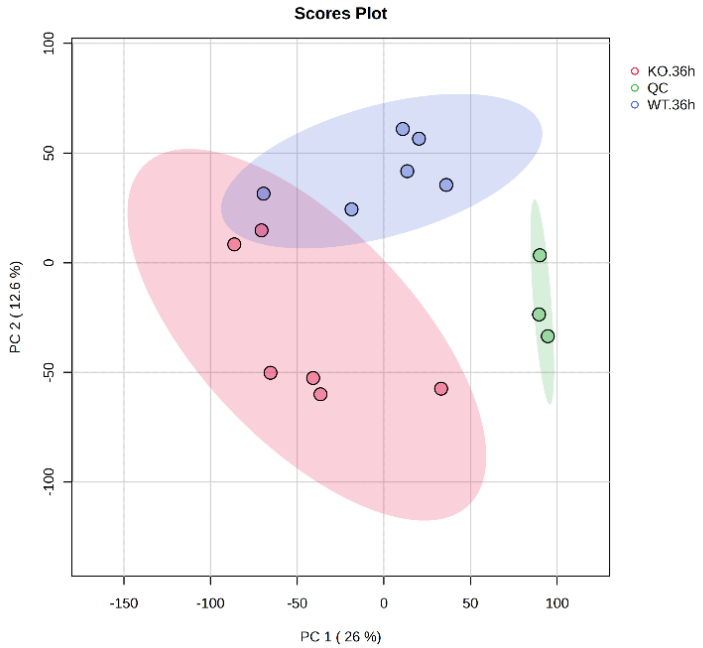

Figure 1 *PCA analysis of the Gene-knockdown (KO), Wild type (WT), and Quality control (QC) groups revealed a cluster within the same sample and a separation between the groups. The QC group, which contained the average amount of metabolite, was separate from another group, which indicates the high reliability of the data.*

6.1.2 Volcano plot

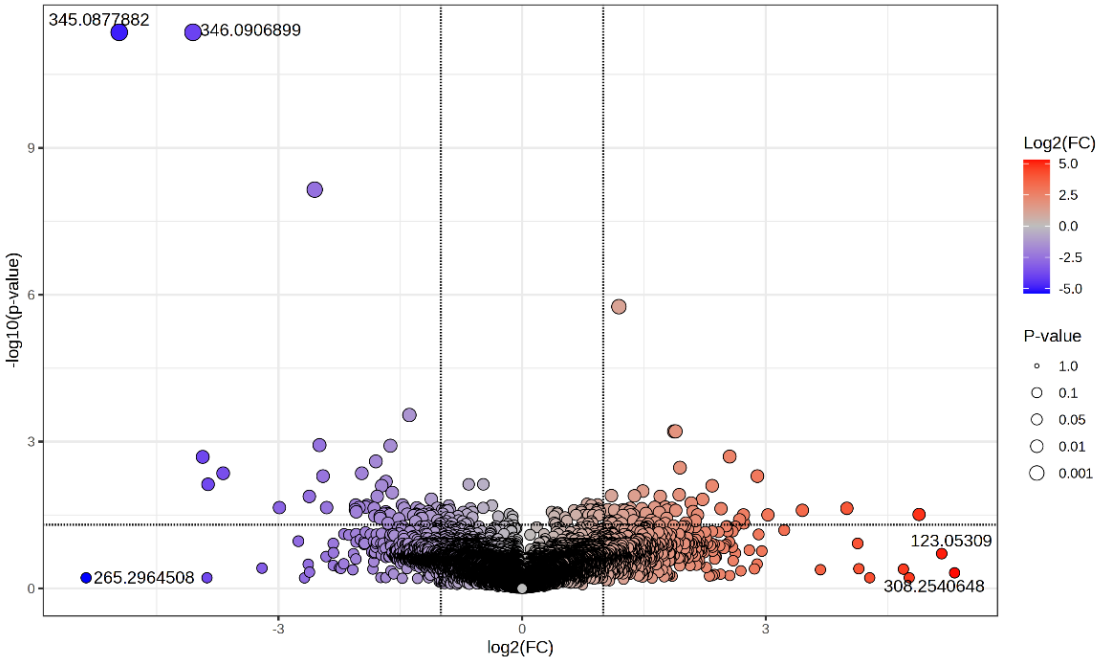

Figure 2 *The volcano plot shows the upregulation and downregulation of each metabolite in a knockdown group compared to the wild-type mouse cell line group. The downregulated compound was labelled in blue, and the line described 2 times fold change (Vertical) and significance level at p-value 0.01 (Horizontal).*

6.1.3 Heatmap

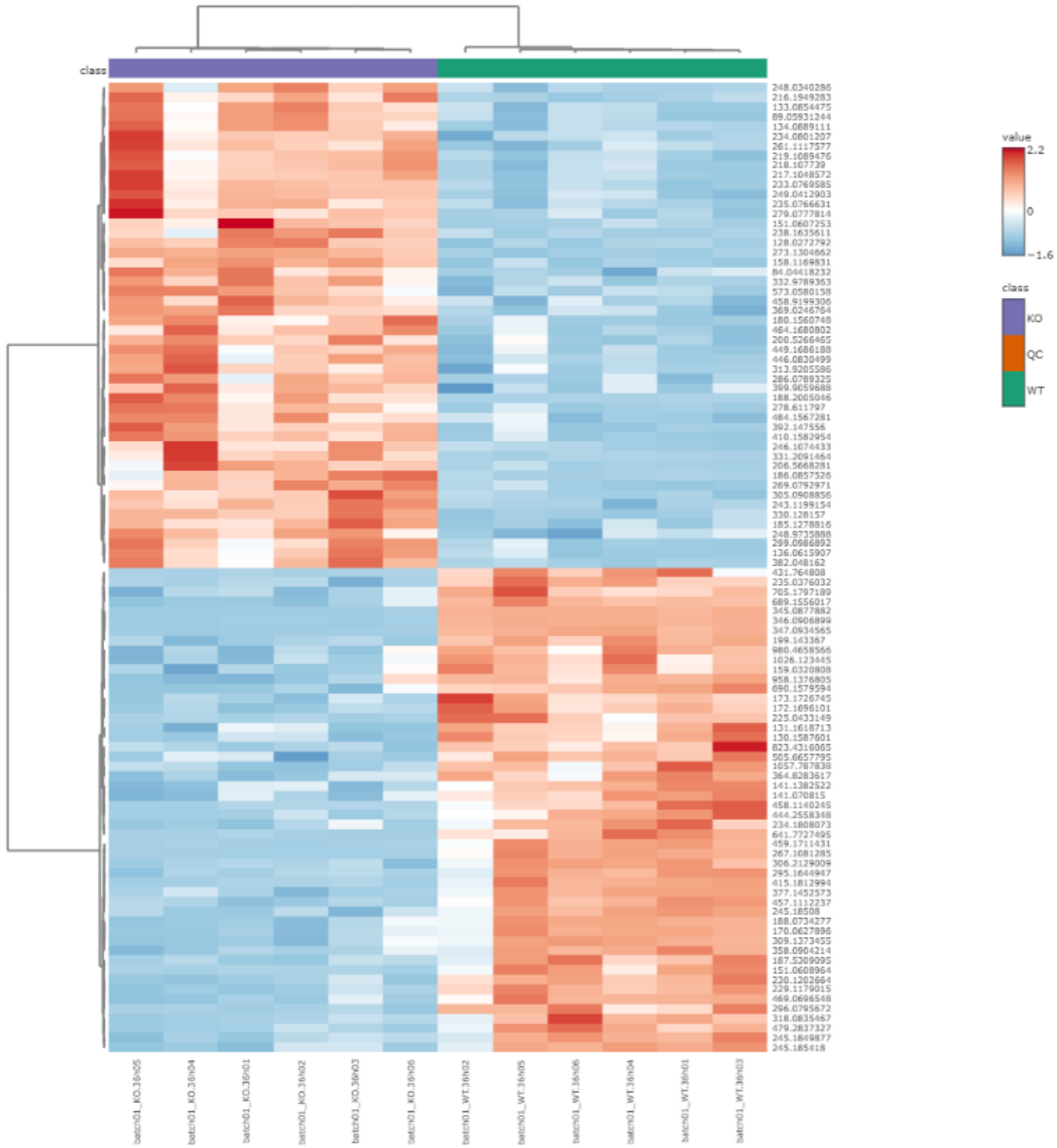

Figure 3 *Heat map analysis of the Gene-knockdown (KO) and wild-type (WT) groups. Different metabolite levels were observed.*

6.2 Identification of compounds

Table 2 *Altered endogenous metabolite, ordered by statistical significance (from highest to lowest). The compound matching the pyruvate, the riboflavin and the glutathione metabolism pathway was highlighted in yellow, pink and green, respectively.*

No	Fold	p-value	UP/DOWN (KO vs WT)	m/z	rt	name	adduct
1	5	2.79E-08	DOWN	229.12	2.05	3-Methoxytyrosine	$M+NH_4$
2	3.1	1.00E-05	DOWN	469.07	1.87	2-(a-Hydroxyethyl)thiamine diphosphate	M+H
3	32.6	3.00E-05	DOWN	345.09	13.88	(1aalpha,2beta,3alpha,11calpha)-1a,2,3,11c-Tetrahydro-6,11-dimethylbenzo[6,7]phenanthro[3,4-b]oxirene-2,3-diol	M+K
4	6.8	0.00007	UP	380.11	5.35	S-Lactoylglutathione	M+H
5	2.5	0.00014	DOWN	245.19	8.64	Leucyl-leucine	M+H
6	14.2	0.00016	DOWN	267.11	6.57	1-(3-Carboxypropyl)-3,7-dimethylxanthine	M+H
7	2.7	0.00019	UP	136.06	7.4	4-Hydroxy-L-threonine	M+H
8	3.7	0.00021	UP	320.08	7.4	5'-Methylthioadenosine	M+Na
9	3.6	0.00024	DOWN	377.15	8.98	Riboflavin	M+H
10	4.9	0.00027	DOWN	457.11	8.66	Flavin mononucleotide	M+H
11	2.6	0.00041	UP	325.04	2.16	Uridine 5'-monophosphate	M+H
12	3.8	0.00065	DOWN	151.06	5.44	1-Methylhypoxanthine	M+H
13	2.1	0.00067	UP	484.20	8.17	Trp Trp Ala	M+Na
14	3.2	0.00074	DOWN	468.31	15.72	LysoPC(14:0)	M+H
15	6.4	0.00082	UP	294.06	1.5	7,8-Dihydroneopterin	M+K
16	4	0.00094	UP	344.09	4.94	L-L-Homoglutathione	M+Na
17	4.3	0.00148	UP	246.11	1.89	Hydroxyprolyl-Asparagine	M+H
18	2.2	0.002	UP	488.30	13.33	Glycocholic Acid	M+Na
19	2.6	0.002	DOWN	302.11	6.24	Asparaginyl-Phenylalanine	M+Na
20	2.3	0.0022	UP	309.23	11.85	N1,N12-Diacetylspermine	M+Na
21	2.4	0.00226	UP	613.16	5	Oxidized glutathione	M+H
22	3.5	0.00242	UP	269.08	2.06	DHAP(6:0)	M+H
23	2.4	0.00245	DOWN	199.14	6.3	Oxilofrin	$M+NH_4$
24	3.8	0.00251	UP	373.22	12.97	13,14-dihydro-16,16-difluoro Prostaglandin J2	M+H
25	2	0.00299	UP	417.16	15.89	Pentosidine	M+K
26	2.4	0.00346	DOWN	188.18	1.39	N1-Acetylspermidine	M+H
27	2	0.0039	UP	130.05	4.99	Pyroglutamic acid	M+H
28	2.4	0.00401	DOWN	823.43	11.24	Loquatifolin A	M+H
29	3	0.00408	UP	424.10	4.8	S-(1,2-Dicarboxyethyl)glutathione	M+H
30	4.7	0.0046	DOWN	166.05	11.66	L-Methionine S-oxide	M+H
31	2.1	0.00474	UP	387.19	14.39	6alpha-Fluoro-11beta,17-dihydroxyprogesterone	M+Na
32	1.7	0.00555	DOWN	273.21	14.06	Tetradecanedioic acid	M+H
33	2.1	0.00597	UP	231.04	5	Indoxyl sulfate	$M+NH_4$
34	10.3	0.00618	UP	158.15	5.97	2-Nonenal	$M+NH_4$
35	3	0.00646	UP	216.20	9.72	11-Dodecenoic acid	$M+NH_4$
36	22.8	0.00721	UP	188.16	5.88	2-amino-decanoic acid	M+H
37	2.1	0.00985	UP	486.28	13.33	Prostaglandin E2 p-acetamidophenyl ester	M+H

6.3 Pathway analysis

Table 3 *The summary of pathway analysis. The top three metabolisms identified were riboflavin, pyruvate and glutathione metabolism*

Pathway Name	Match Status	p	-log(p)	Holm p	FDR	Impact
Riboflavin metabolism	2/4	3.37E-04	3.4725	0.026954	0.026954	1
Pyruvate metabolism	2/23	0.013073	1.8836	1	0.50958	0.22372
Glutathione metabolism	2/28	0.019109	1.7188	1	0.50958	0.03407
Citrate cycle (TCA cycle)	1/20	0.14682	0.83321	1	1	0.06592
Glycolysis / Gluconeogenesis	1/26	0.18685	0.72852	1	1	0.08574
Folate biosynthesis	1/27	0.19335	0.71366	1	1	0
Lipoic acid metabolism	1/28	0.1998	0.69941	1	1	0.03561
Cysteine and methionine metabolism	1/33	0.23136	0.63571	1	1	0.02089
Glycerophospholipid metabolism	1/36	0.24975	0.6025	1	1	0.01736
Pyrimidine metabolism	1/39	0.26773	0.57231	1	1	0.08327
Primary bile acid biosynthesis	1/46	0.30816	0.51123	1	1	0.02285
Metabolism of xenobiotics by cytochrome P450	1/64	0.40291	0.39479	1	1	0

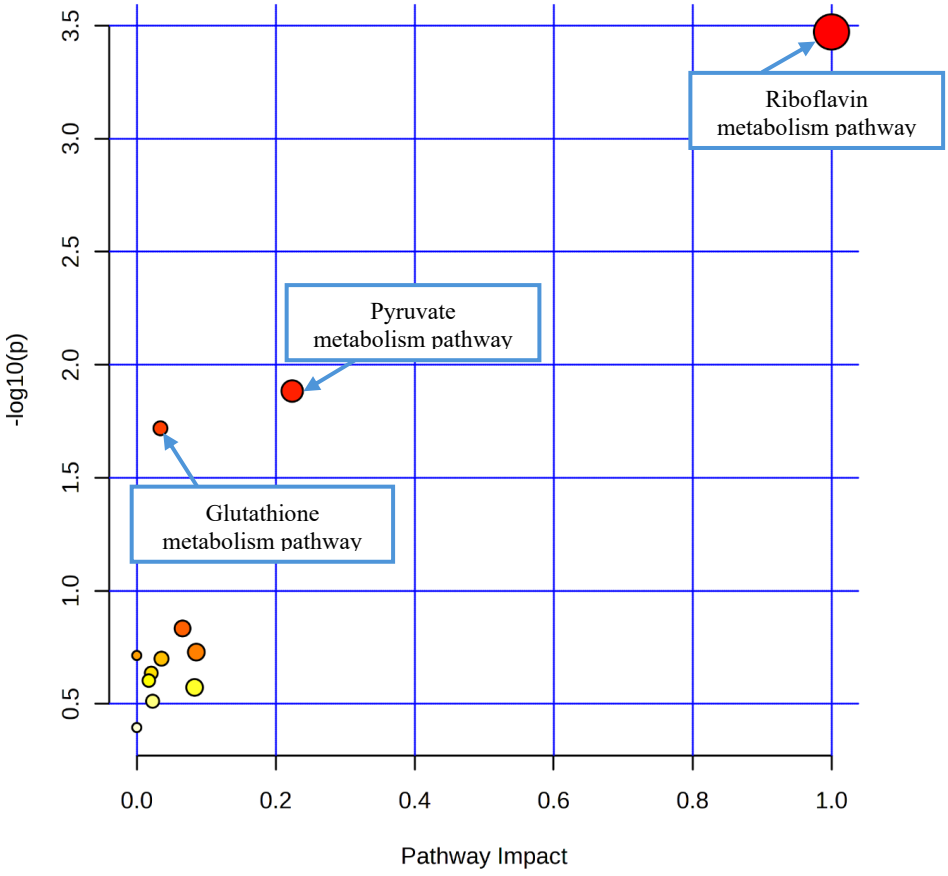

Figure 4 *In the Pathway analysis plot, the pathway impact (x-axis) was plotted against a significant level, as -log$_{10}$(p) (y-axis). The three most significant pathways were Riboflavin, Pyruvate and Glutathione metabolisms*

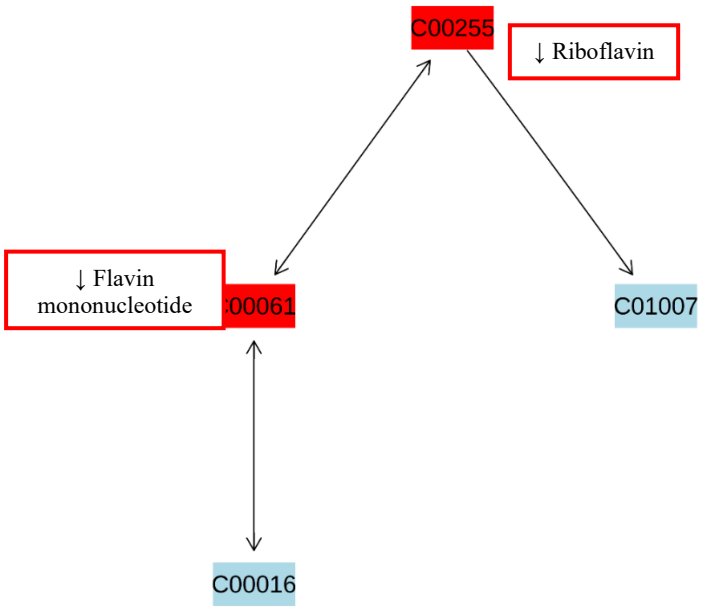

Figure 5 The Riboflavin metabolism pathway, the downregulation of Riboflavin (Vitamin B2) and flavin mononucleotide were observed.

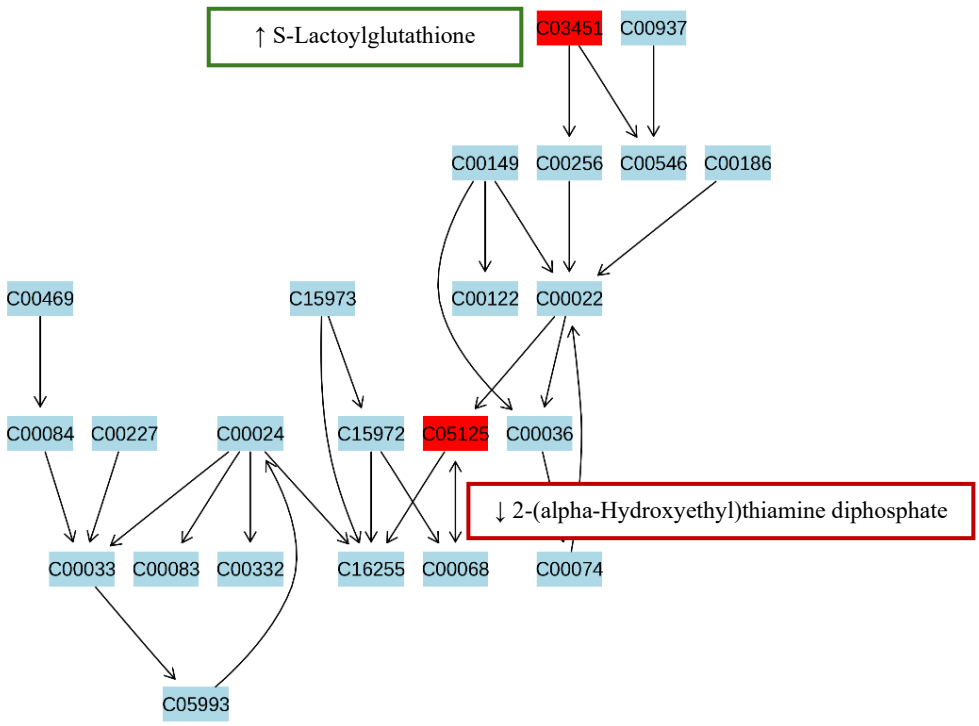

Figure 6 The Pyruvate metabolism pathway, the downregulation of 2-(alpha-Hydroxyethyl)thiamine diphosphate and upregulation of S-Lactoylglutathione were observed

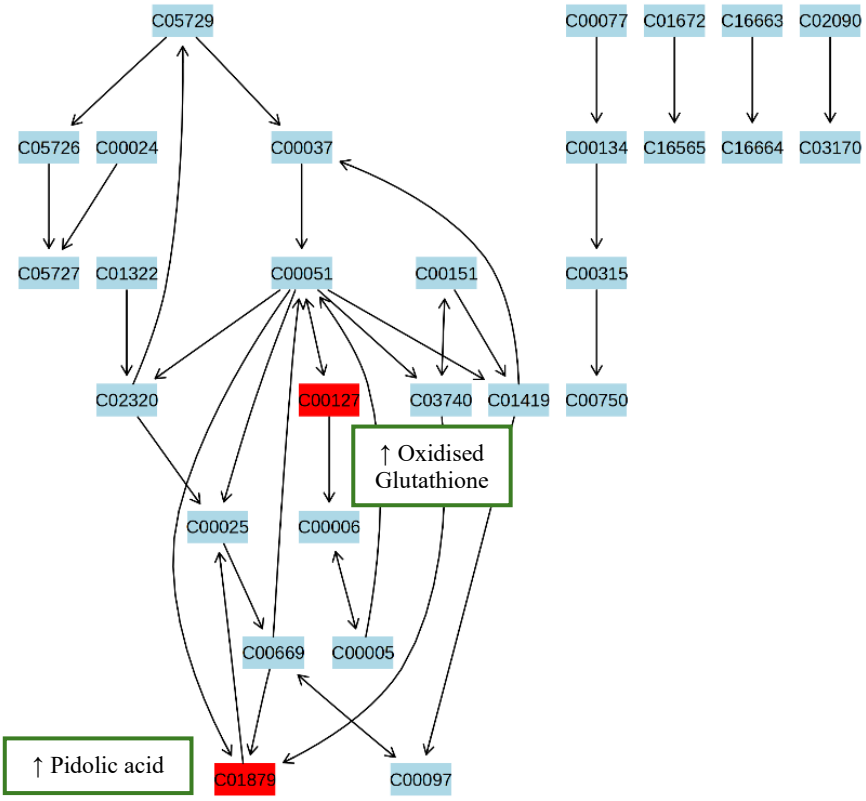

Figure 7 The Glutathione metabolism pathway. The upregulation of Oxidised glutathione and Pidolic acid were observed.