Coding with R
MetaboAnalyst

07 DECEMBER 2022

INGVILD COMFORT HVINDEN

DPHIL, MCCULLAGH GROUP

What is R?

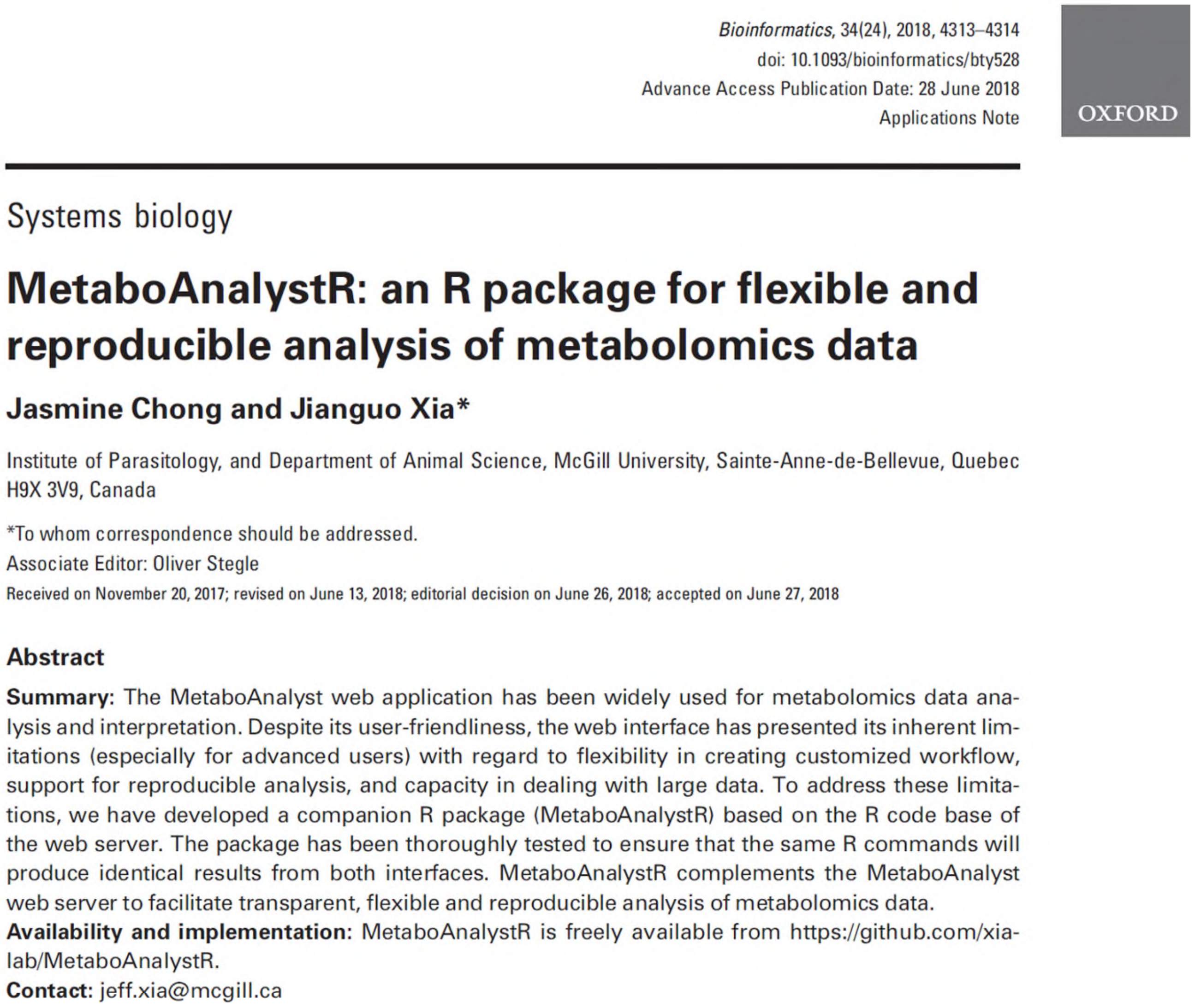
R is a programming language available for free on https://www.r-project.org/

R can be run via RStudio

Both R and RStudio can be used on Windows, macOS and Linux

It’s used mainly for data mining, statistical analysis and data visualization

There are a range of ‘packages’ that provide additional tools for statistical analysis, visualization and more.



Bioinformatics, 34(24), 2018, 4313–4314
doi: 10.1093/bioinformatics/bty528
Advance Access Publication Date: 28 June 2018
Applications Note

Systems biology

MetaboAnalystR: an R package for flexible and reproducible analysis of metabolomics data

Jasmine Chong and Jianguo Xia*

Institute of Parasitology, and Department of Animal Science, McGill University, Sainte-Anne-de-Bellevue, Quebec H9X 3V9, Canada

*To whom correspondence should be addressed.

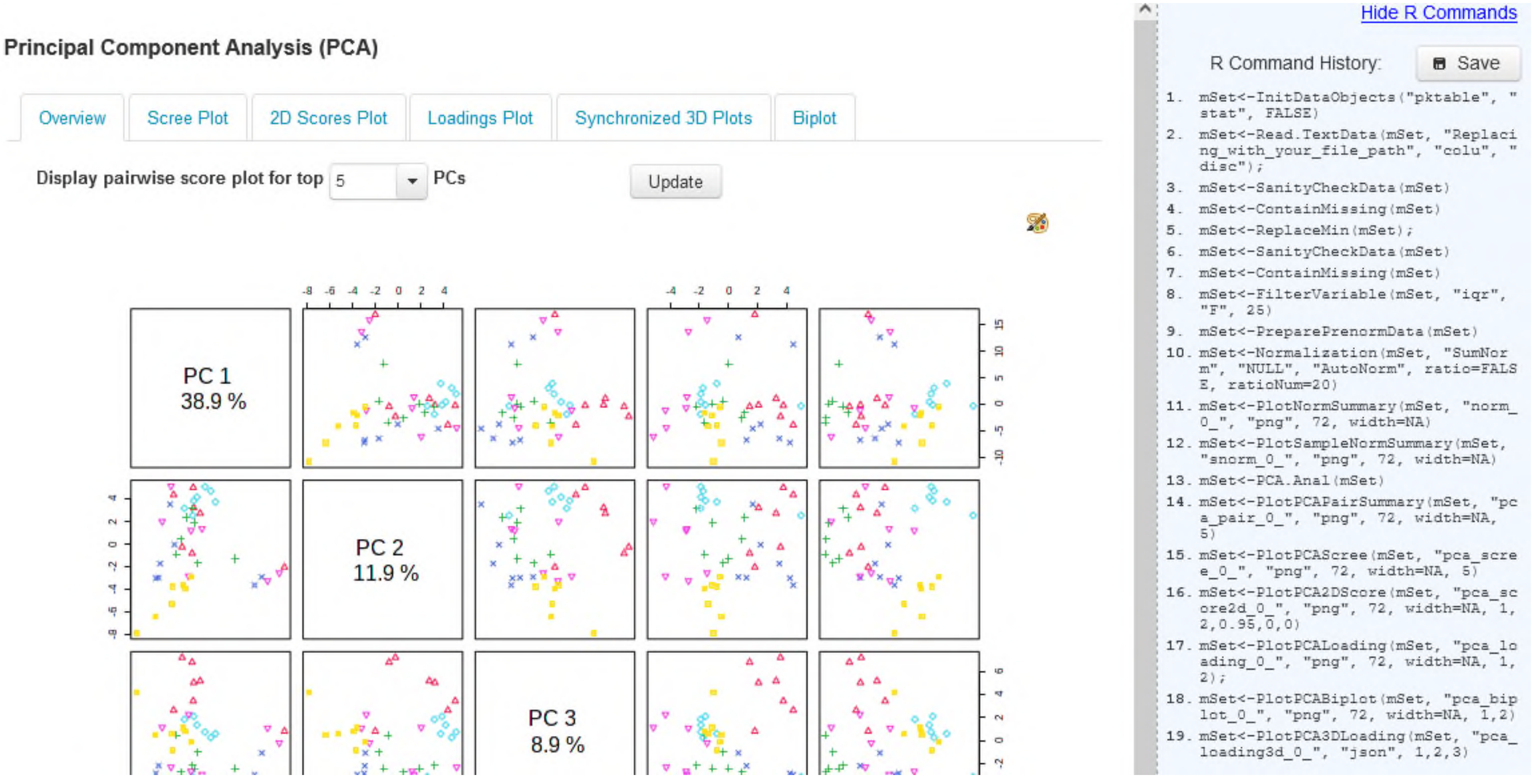
Associate Editor: Oliver Stegle
Received on November 20, 2017; revised on June 13, 2018; editorial decision on June 26, 2018; accepted on June 27, 2018

Abstract

Summary: The MetaboAnalyst web application has been widely used for metabolomics data analysis and interpretation. Despite its user-friendliness, the web interface has presented its inherent limitations (especially for advanced users) with regard to flexibility in creating customized workflow, support for reproducible analysis, and capacity in dealing with large data. To address these limitations, we have developed a companion R package (MetaboAnalystR) based on the R code base of the web server. The package has been thoroughly tested to ensure that the same R commands will produce identical results from both interfaces. MetaboAnalystR complements the MetaboAnalyst web server to facilitate transparent, flexible and reproducible analysis of metabolomics data.
Availability and implementation: MetaboAnalystR is freely available from https://github.com/xia-lab/MetaboAnalystR.
Contact: jeff.xia@mcgill.ca

MetaboAnalyst is run with R code

WEBSITE TUTORIAL:
HTTPS://WWW.METABOANALYST.CA/DOCS/RTUTORIAL.XHTML



Principal Component Analysis (PCA)

```
1.  mSet<-InitDataObjects("pktable", "
    stat", FALSE)
2.  mSet<-Read.TextData(mSet, "Replaci
    ng_with_your_file_path", "colu", "
    disc");
3.  mSet<-SanityCheckData(mSet)
4.  mSet<-ContainMissing(mSet)
5.  mSet<-ReplaceMin(mSet);
6.  mSet<-SanityCheckData(mSet)
7.  mSet<-ContainMissing(mSet)
8.  mSet<-FilterVariable(mSet, "iqr",
    "F", 25)
9.  mSet<-PreparePrenormData(mSet)
10. mSet<-Normalization(mSet, "SumNor
    m", "NULL", "AutoNorm", ratio=FALS
    E, ratioNum=20)
11. mSet<-PlotNormSummary(mSet, "norm_
    0_", "png", 72, width=NA)
12. mSet<-PlotSampleNormSummary(mSet,
    "snorm_0_", "png", 72, width=NA)
13. mSet<-PCA.Anal(mSet)
14. mSet<-PlotPCAPairSummary(mSet, "pc
    a_pair_0_", "png", 72, width=NA,
    5)
15. mSet<-PlotPCAScree(mSet, "pca_scre
    e_0_", "png", 72, width=NA, 5)
16. mSet<-PlotPCA2DScore(mSet, "pca_sc
    ore2d_0_", "png", 72, width=NA, 1,
    2,0.95,0,0)
17. mSet<-PlotPCALoading(mSet, "pca_lo
    ading_0_", "png", 72, width=NA, 1,
    2);
18. mSet<-PlotPCABiplot(mSet, "pca_bip
    lot_0_", "png", 72, width=NA, 1,2)
19. mSet<-PlotPCA3DLoading(mSet, "pca_
    loading3d_0_", "json", 1,2,3)
```

MetaboAnalyst is run with R code

WEBSITE TUTORIAL: [HTTPS://WWW.METABOANALYST.CA/DOCS/RTUTORIAL.XHTML](https://www.metaboanalyst.ca/docs/RTutorial.xhtml)

Challenges in metabolomics data analysis

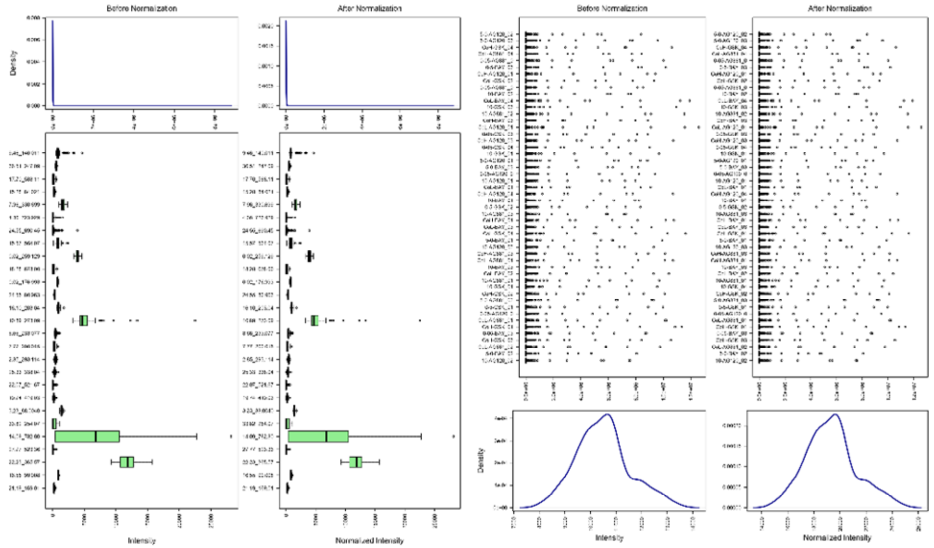
- Metabolomics can generate many data sets
- Data sets often contain several different experimental groups
- Leads to cumbersome and time-consuming data processing

Solution: R

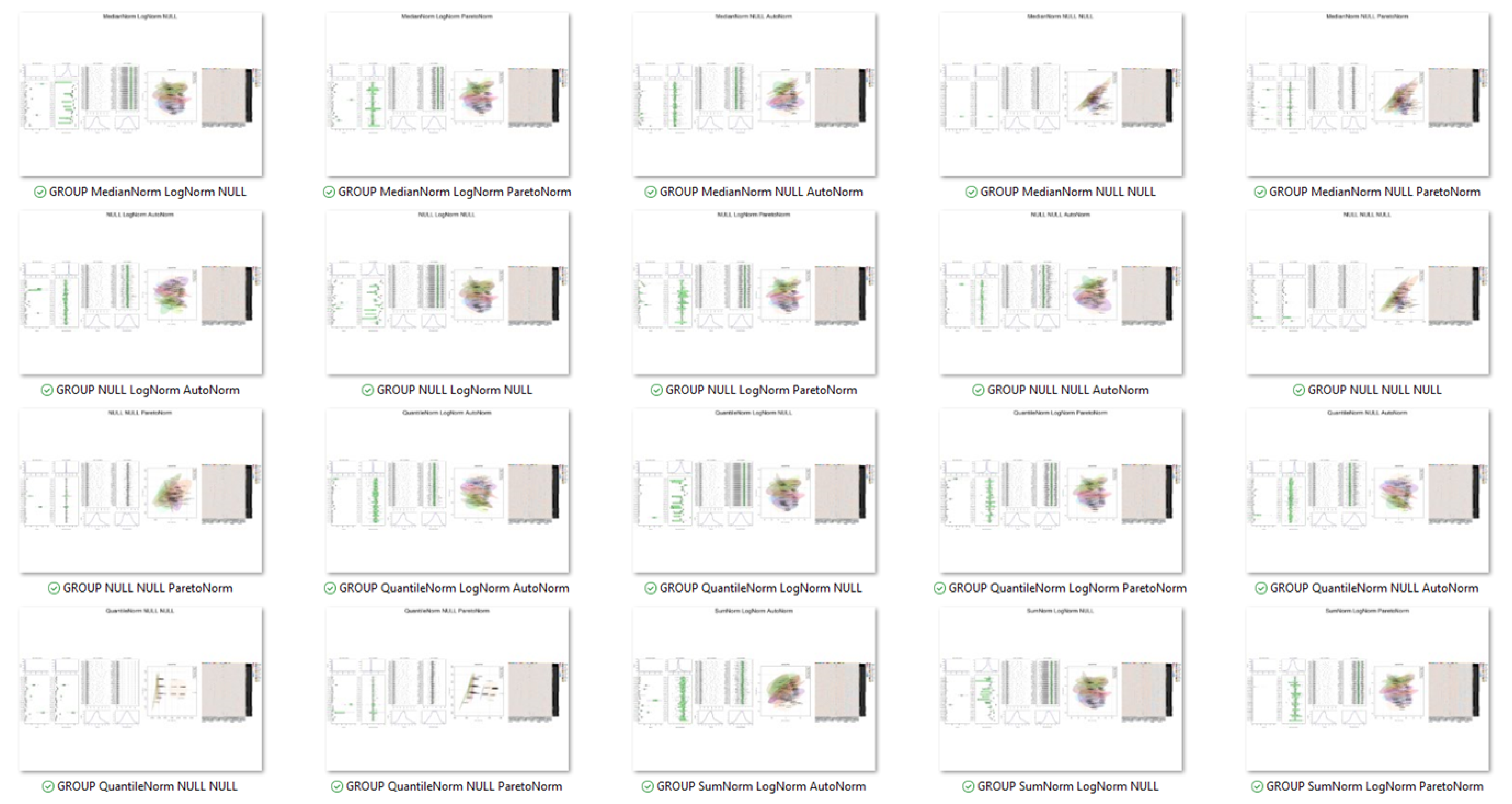
Save time by automating repetitive tasks

- Figuring out normalisation, transformation and scaling parameters and identifying outliers
- Calculate mean and standard deviation for all features
- Plotting tens, hundreds or thousands of features
- Carrying out multiple binary comparisons
  - Fold change
  - Significance tests
  - Functional analysis

Choosing normalisation, scaling and transformation parameters

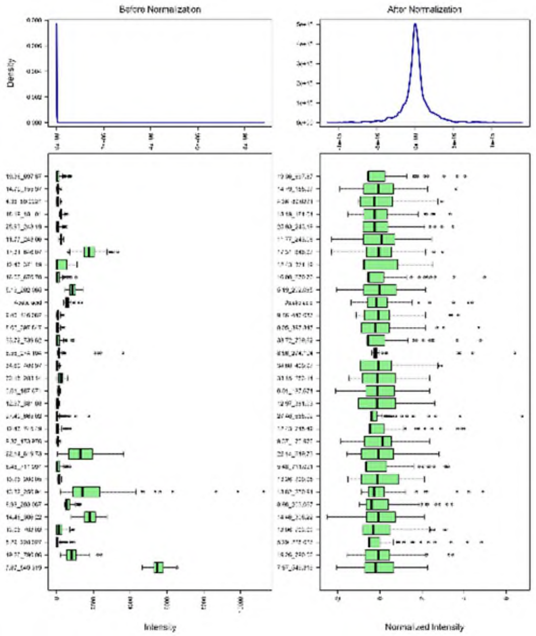


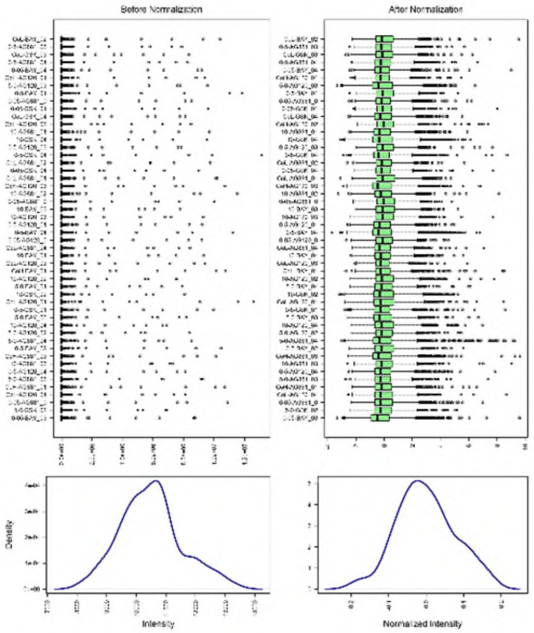
No normalisation, scaling
or transformation applied

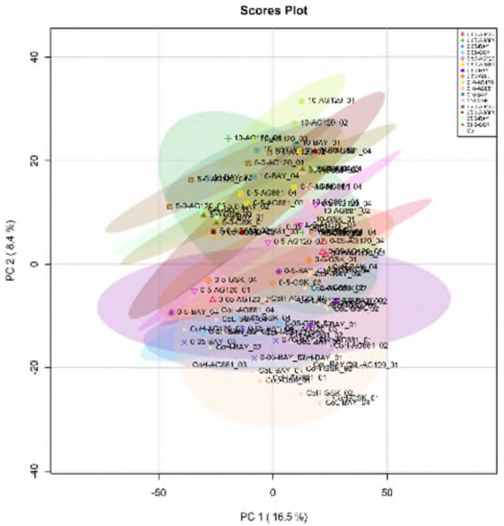


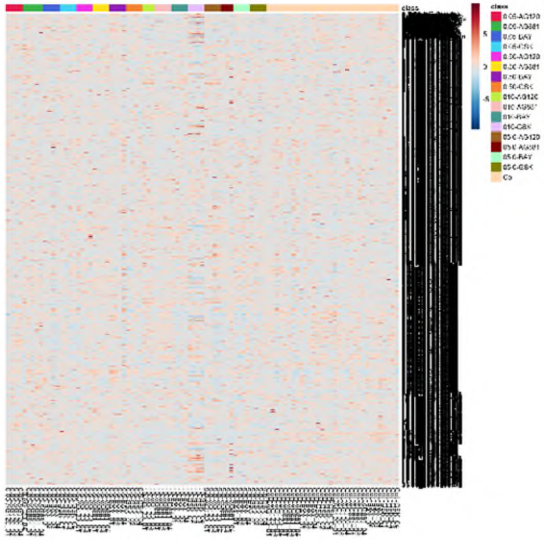
✓ GROUP MedianNorm LogNorm NULL
✓ GROUP MedianNorm LogNorm ParetoNorm
✓ GROUP MedianNorm NULL AutoNorm
✓ GROUP MedianNorm NULL NULL
✓ GROUP MedianNorm NULL ParetoNorm
✓ GROUP NULL LogNorm AutoNorm
✓ GROUP NULL LogNorm NULL
✓ GROUP NULL LogNorm ParetoNorm
✓ GROUP NULL NULL AutoNorm
✓ GROUP NULL NULL NULL
✓ GROUP NULL NULL ParetoNorm
✓ GROUP QuantileNorm LogNorm AutoNorm
✓ GROUP QuantileNorm LogNorm NULL
✓ GROUP QuantileNorm LogNorm ParetoNorm
✓ GROUP QuantileNorm NULL AutoNorm
✓ GROUP QuantileNorm NULL NULL
✓ GROUP QuantileNorm NULL ParetoNorm
✓ GROUP SumNorm LogNorm AutoNorm
✓ GROUP SumNorm LogNorm NULL
✓ GROUP SumNorm LogNorm ParetoNorm

Choosing normalisation, scaling and transformation parameters



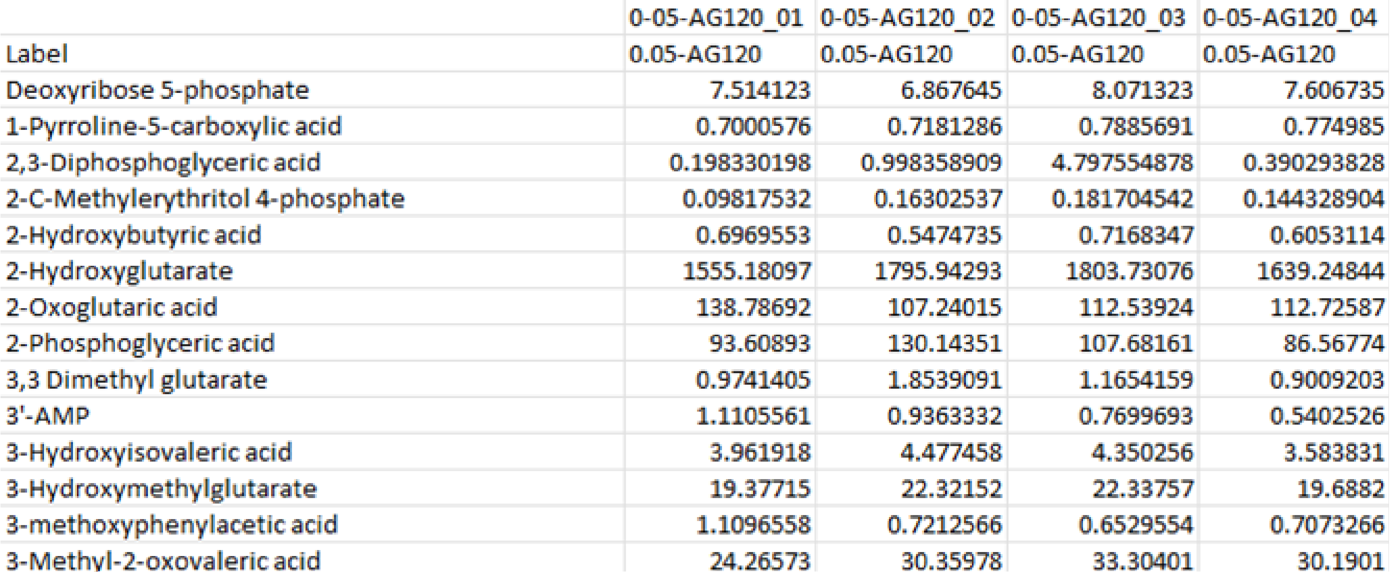




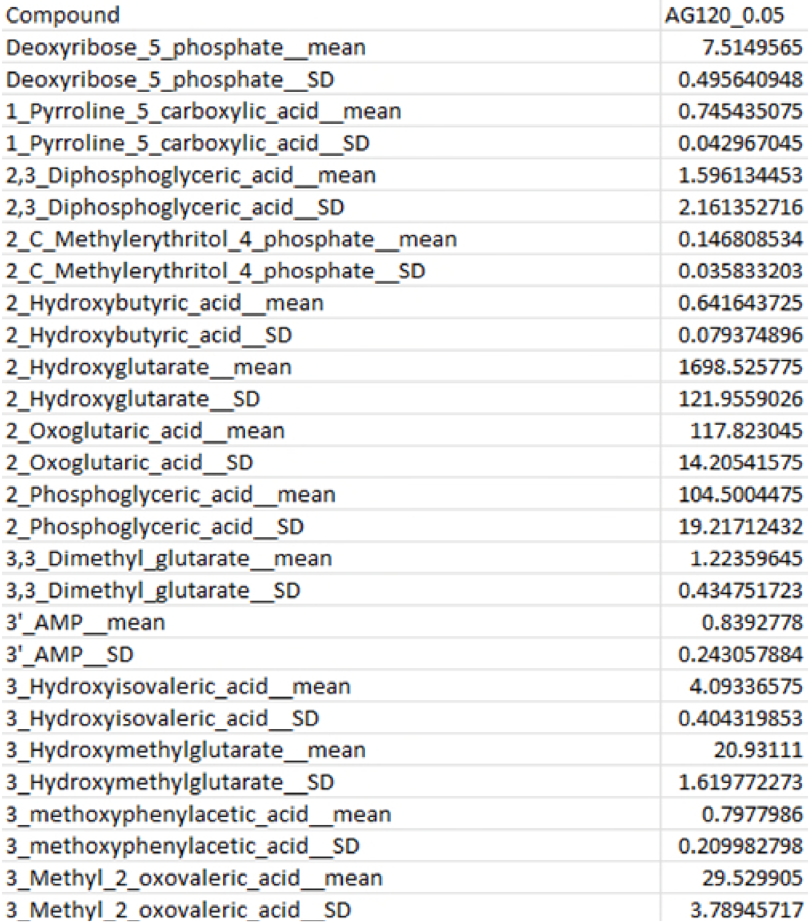


Median normalisation
Auto scaling

Calculating mean and standard deviation

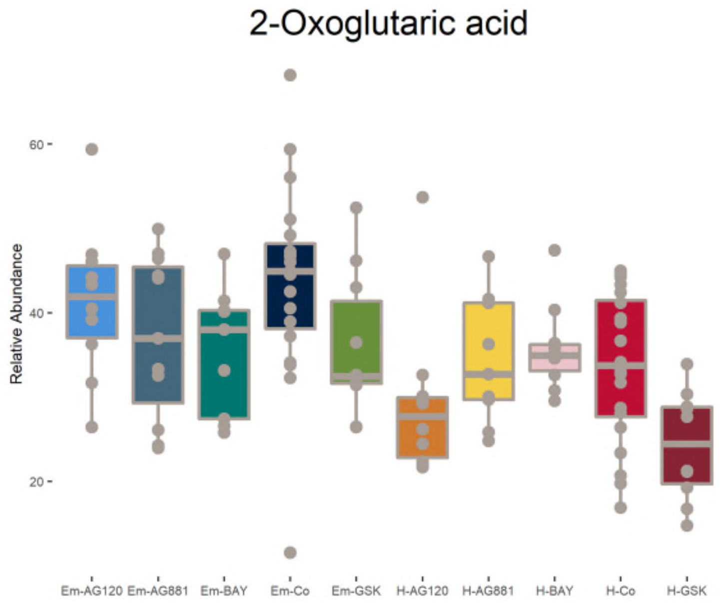


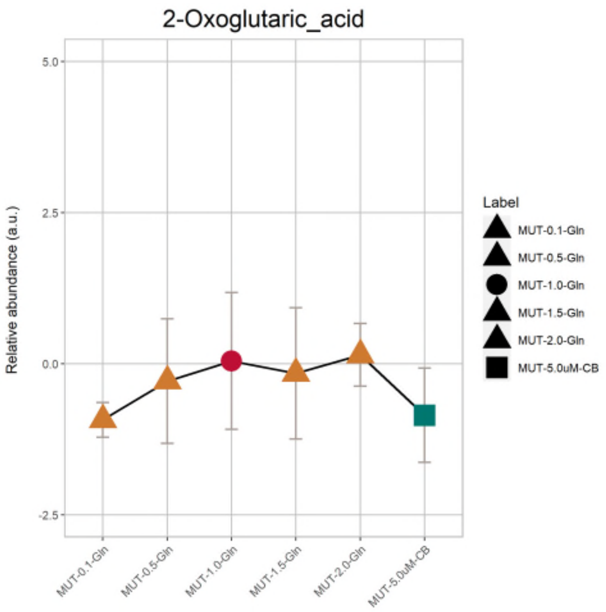
	0-05-AG120_01	0-05-AG120_02	0-05-AG120_03	0-05-AG120_04
Label	0.05-AG120	0.05-AG120	0.05-AG120	0.05-AG120
Deoxyribose 5-phosphate	7.514123	6.867645	8.071323	7.606735
1-Pyrroline-5-carboxylic acid	0.7000576	0.7181286	0.7885691	0.774985
2,3-Diphosphoglyceric acid	0.198330198	0.998358909	4.797554878	0.390293828
2-C-Methylerythritol 4-phosphate	0.09817532	0.16302537	0.181704542	0.144328904
2-Hydroxybutyric acid	0.6969553	0.5474735	0.7168347	0.6053114
2-Hydroxyglutarate	1555.18097	1795.94293	1803.73076	1639.24844
2-Oxoglutaric acid	138.78692	107.24015	112.53924	112.72587
2-Phosphoglyceric acid	93.60893	130.14351	107.68161	86.56774
3,3 Dimethyl glutarate	0.9741405	1.8539091	1.1654159	0.9009203
3'-AMP	1.1105561	0.9363332	0.7699693	0.5402526
3-Hydroxyisovaleric acid	3.961918	4.477458	4.350256	3.583831
3-Hydroxymethylglutarate	19.37715	22.32152	22.33757	19.6882
3-methoxyphenylacetic acid	1.1096558	0.7212566	0.6529554	0.7073266
3-Methyl-2-oxovaleric acid	24.26573	30.35978	33.30401	30.1901



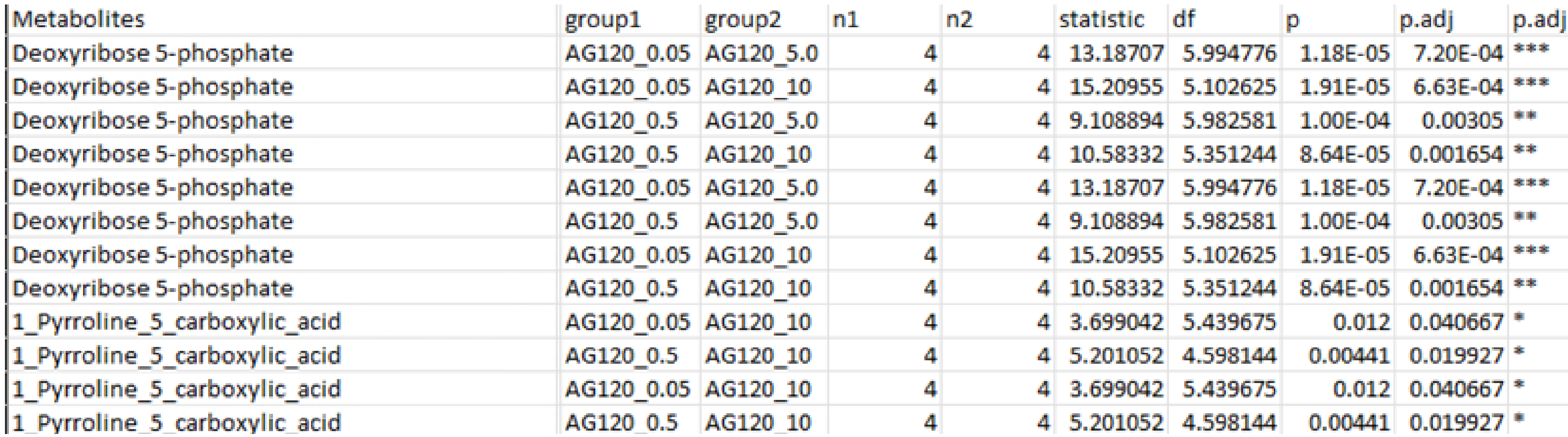
Compound	AG120_0.05
Deoxyribose_5_phosphate__mean	7.5149565
Deoxyribose_5_phosphate__SD	0.495640948
1_Pyrroline_5_carboxylic_acid__mean	0.745435075
1_Pyrroline_5_carboxylic_acid__SD	0.042967045
2,3_Diphosphoglyceric_acid__mean	1.596134453
2,3_Diphosphoglyceric_acid__SD	2.161352716
2_C_Methylerythritol_4_phosphate__mean	0.146808534
2_C_Methylerythritol_4_phosphate__SD	0.035833203
2_Hydroxybutyric_acid__mean	0.641643725
2_Hydroxybutyric_acid__SD	0.079374896
2_Hydroxyglutarate__mean	1698.525775
2_Hydroxyglutarate__SD	121.9559026
2_Oxoglutaric_acid__mean	117.823045
2_Oxoglutaric_acid__SD	14.20541575
2_Phosphoglyceric_acid__mean	104.5004475
2_Phosphoglyceric_acid__SD	19.21712432
3,3_Dimethyl_glutarate__mean	1.22359645
3,3_Dimethyl_glutarate__SD	0.434751723
3'_AMP__mean	0.8392778
3'_AMP__SD	0.243057884
3_Hydroxyisovaleric_acid__mean	4.09336575
3_Hydroxyisovaleric_acid__SD	0.404319853
3_Hydroxymethylglutarate__mean	20.93111
3_Hydroxymethylglutarate__SD	1.619772273
3_methoxyphenylacetic_acid__mean	0.7977986
3_methoxyphenylacetic_acid__SD	0.209982798
3_Methyl_2_oxovaleric_acid__mean	29.529905
3_Methyl_2_oxovaleric_acid__SD	3.78945717

Efficient plotting of features and metabolites

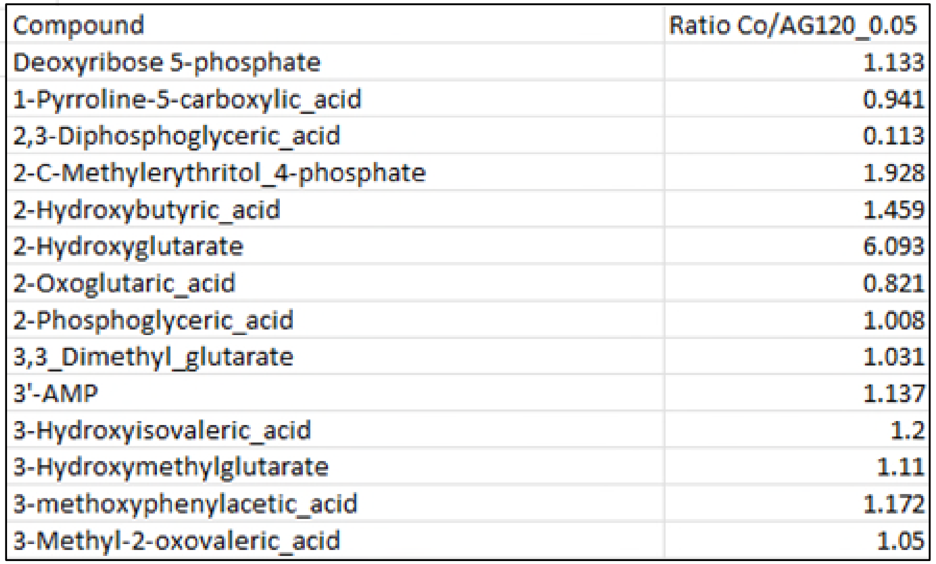




Significance tests and fold change calculations



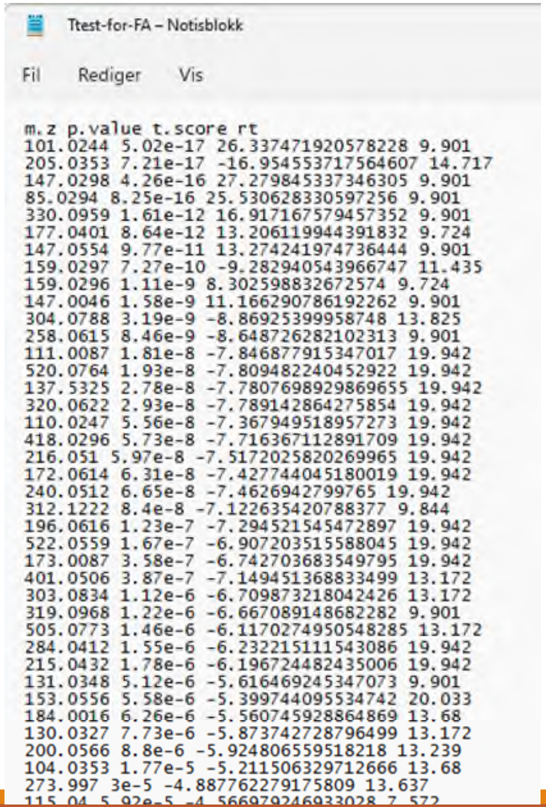
Metabolites	group1	group2	n1	n2	statistic	df	p	p.adj	p.adj.signif
Deoxyribose 5-phosphate	AG120_0.05	AG120_5.0	4	4	13.18707	5.994776	1.18E-05	7.20E-04	***
Deoxyribose 5-phosphate	AG120_0.05	AG120_10	4	4	15.20955	5.102625	1.91E-05	6.63E-04	***
Deoxyribose 5-phosphate	AG120_0.5	AG120_5.0	4	4	9.108894	5.982581	1.00E-04	0.00305	**
Deoxyribose 5-phosphate	AG120_0.5	AG120_10	4	4	10.58332	5.351244	8.64E-05	0.001654	**
Deoxyribose 5-phosphate	AG120_0.05	AG120_5.0	4	4	13.18707	5.994776	1.18E-05	7.20E-04	***
Deoxyribose 5-phosphate	AG120_0.5	AG120_5.0	4	4	9.108894	5.982581	1.00E-04	0.00305	**
Deoxyribose 5-phosphate	AG120_0.05	AG120_10	4	4	15.20955	5.102625	1.91E-05	6.63E-04	***
Deoxyribose 5-phosphate	AG120_0.5	AG120_10	4	4	10.58332	5.351244	8.64E-05	0.001654	**
1_Pyrroline_5_carboxylic_acid	AG120_0.05	AG120_10	4	4	3.699042	5.439675	0.012	0.040667	*
1_Pyrroline_5_carboxylic_acid	AG120_0.5	AG120_10	4	4	5.201052	4.598144	0.00441	0.019927	*
1_Pyrroline_5_carboxylic_acid	AG120_0.05	AG120_10	4	4	3.699042	5.439675	0.012	0.040667	*
1_Pyrroline_5_carboxylic_acid	AG120_0.5	AG120_10	4	4	5.201052	4.598144	0.00441	0.019927	*



Compound	Ratio Co/AG120_0.05
Deoxyribose 5-phosphate	1.133
1-Pyrroline-5-carboxylic_acid	0.941
2,3-Diphosphoglyceric_acid	0.113
2-C-Methylerythritol_4-phosphate	1.928
2-Hydroxybutyric_acid	1.459
2-Hydroxyglutarate	6.093
2-Oxoglutaric_acid	0.821
2-Phosphoglyceric_acid	1.008
3,3_Dimethyl_glutarate	1.031
3'-AMP	1.137
3-Hydroxyisovaleric_acid	1.2
3-Hydroxymethylglutarate	1.11
3-methoxyphenylacetic_acid	1.172
3-Methyl-2-oxovaleric_acid	1.05

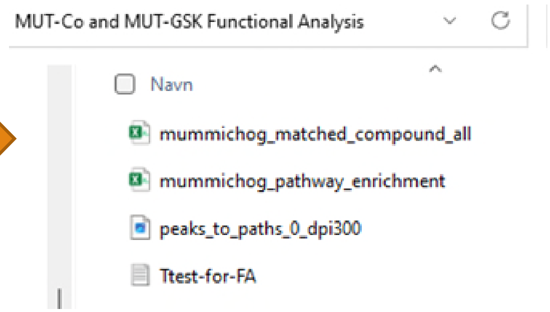
Functional (untargeted) analysis

Instead of manually preparing t-tests for all binary comparisons and uploading to metaboanalyst.ca, the code calculates and performs the analysis automatically



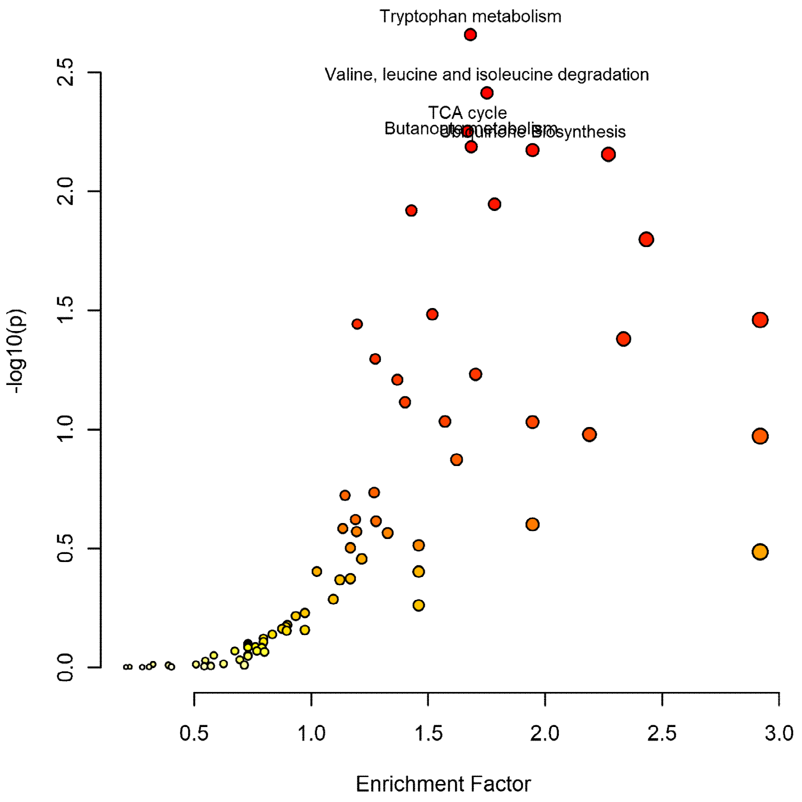
```
m.z p.value t.score rt
101.0244 5.02e-17 26.337471920578228 9.901
205.0353 7.21e-17 -16.954553717564607 14.717
147.0298 4.26e-16 27.279845337346305 9.901
85.0294 8.25e-16 25.530628330597256 9.901
330.0959 1.61e-12 16.917167579457352 9.901
177.0401 8.64e-12 13.206119944391832 9.724
147.0554 9.77e-11 13.274241974736444 9.901
159.0297 7.27e-10 -9.282940543966747 11.435
159.0296 1.11e-9 8.302598832672574 9.724
147.0046 1.58e-9 11.166290786192262 9.901
304.0788 3.19e-9 -8.86925399958748 13.825
258.0615 8.46e-9 -8.648726282102313 9.901
111.0087 1.81e-8 -7.846877915347017 19.942
520.0764 1.93e-8 -7.809482240452922 19.942
137.5325 2.78e-8 -7.7807698929869655 19.942
320.0622 2.93e-8 -7.789142864275854 19.942
110.0247 5.56e-8 -7.367949518957273 19.942
418.0296 5.73e-8 -7.716367112891709 19.942
216.051 5.97e-8 -7.5172025820269965 19.942
172.0614 6.31e-8 -7.427744045180019 19.942
240.0512 6.65e-8 -7.4626942799765 19.942
312.1222 8.4e-8 -7.122635420788377 9.844
196.0616 1.23e-7 -7.294521545472897 19.942
522.0559 1.67e-7 -6.907203515588045 19.942
173.0087 3.58e-7 -6.742703683549795 19.942
401.0506 3.87e-7 -7.149451368833499 13.172
303.0834 1.12e-6 -6.709873218042426 13.172
319.0968 1.22e-6 -6.667089148682282 9.901
505.0773 1.46e-6 -6.1170274950548285 13.172
284.0412 1.55e-6 -6.232215111543086 19.942
215.0432 1.78e-6 -6.196724482435006 19.942
131.0348 5.12e-6 -5.616469245347073 9.901
153.0556 5.58e-6 -5.399744095534742 20.033
184.0016 6.26e-6 -5.560745928864869 13.68
130.0327 7.73e-6 -5.873742728796499 13.172
200.0566 8.8e-6 -5.924806559518218 13.239
104.0353 1.77e-5 -5.211506329712666 13.68
273.997 3e-5 -4.887762279175809 13.637
115.04 5.92e-5 -4.566979246933028 7.572
```

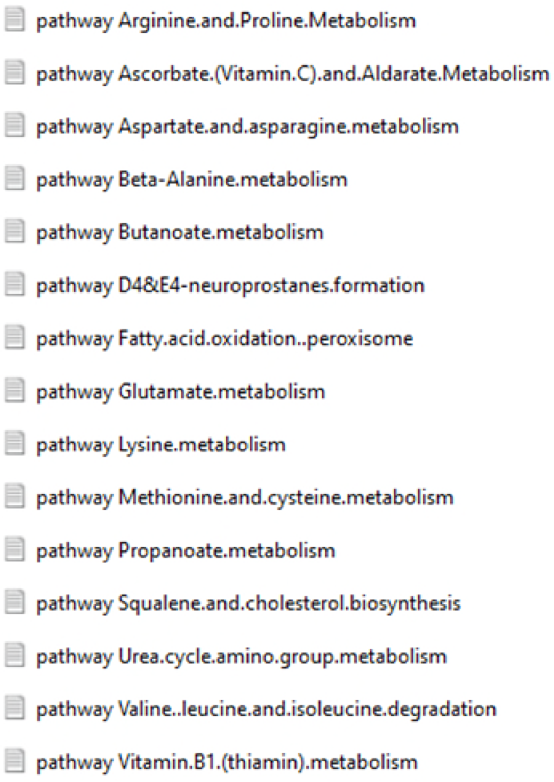
- MUT-Co and MUT-AG120 Functional Analysis
- MUT-Co and MUT-AG881 Functional Analysis
- MUT-Co and MUT-BAY Functional Analysis
- MUT-Co and MUT-GSK Functional Analysis
- MUT-Co and WT-Co Functional Analysis
- MUT-Co and WT-Co Functional Analysis no 2-HG
- WT-Co and WT-AG120 Functional Analysis
- WT-Co and WT-AG881 Functional Analysis
- WT-Co and WT-BAY Functional Analysis
- WT-Co and WT-GSK Functional Analysis

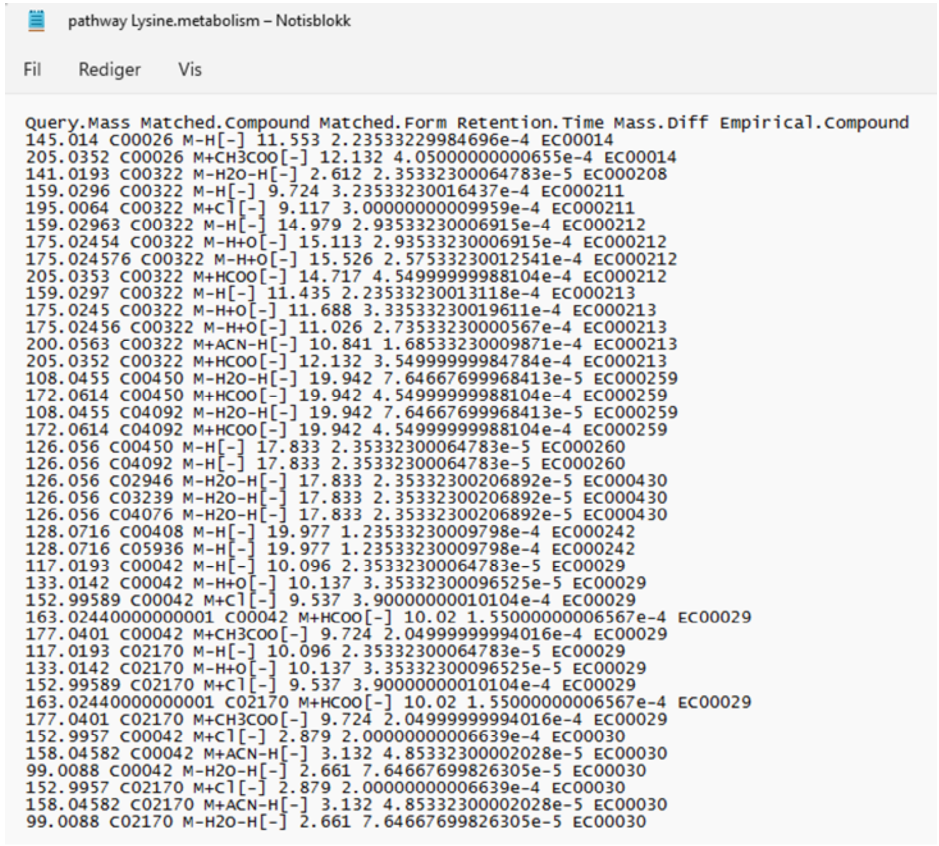


- mummichog_matched_compound_all
- mummichog_pathway_enrichment
- peaks_to_paths_0_dpi300
- Ttest-for-FA

Functional (untargeted) analysis




pathway Arginine.and.Proline.Metabolism
pathway Ascorbate.(Vitamin.C).and.Aldarate.Metabolism
pathway Aspartate.and.asparagine.metabolism
pathway Beta-Alanine.metabolism
pathway Butanoate.metabolism
pathway D4&E4-neuroprostanes.formation
pathway Fatty.acid.oxidation..peroxisome
pathway Glutamate.metabolism
pathway Lysine.metabolism
pathway Methionine.and.cysteine.metabolism
pathway Propanoate.metabolism
pathway Squalene.and.cholesterol.biosynthesis
pathway Urea.cycle.amino.group.metabolism
pathway Valine..leucine.and.isoleucine.degradation
pathway Vitamin.B1.(thiamin).metabolism



```
Query.Mass Matched.Compound Matched.Form Retention.Time Mass.Diff Empirical.Compound
145.014 C00026 M-H[-] 11.553 2.23533229984696e-4 EC00014
205.0352 C00026 M+CH3COO[-] 12.132 4.05000000000655e-4 EC00014
141.0193 C00322 M-H2O-H[-] 2.612 2.35332300064783e-5 EC000208
159.0296 C00322 M-H[-] 9.724 3.23533230016437e-4 EC000211
195.0064 C00322 M+Cl[-] 9.117 3.00000000009959e-4 EC000211
159.02963 C00322 M-H[-] 14.979 2.93533230006915e-4 EC000212
175.02454 C00322 M-H+O[-] 15.113 2.93533230006915e-4 EC000212
175.024576 C00322 M-H+O[-] 15.526 2.57533230012541e-4 EC000212
205.0353 C00322 M+HCOO[-] 14.717 4.54999999988104e-4 EC000212
159.0297 C00322 M-H[-] 11.435 2.23533230013118e-4 EC000213
175.0245 C00322 M-H+O[-] 11.688 3.33533230019611e-4 EC000213
175.02456 C00322 M-H+O[-] 11.026 2.73533230000567e-4 EC000213
200.0563 C00322 M+ACN-H[-] 10.841 1.68533230009871e-4 EC000213
205.0352 C00322 M+HCOO[-] 12.132 3.54999999984784e-4 EC000213
108.0455 C00450 M-H2O-H[-] 19.942 7.64667699968413e-5 EC000259
172.0614 C00450 M+HCOO[-] 19.942 4.54999999988104e-4 EC000259
108.0455 C04092 M-H2O-H[-] 19.942 7.64667699968413e-5 EC000259
172.0614 C04092 M+HCOO[-] 19.942 4.54999999988104e-4 EC000259
126.056 C00450 M-H[-] 17.833 2.35332300064783e-5 EC000260
126.056 C04092 M-H[-] 17.833 2.35332300064783e-5 EC000260
126.056 C02946 M-H2O-H[-] 17.833 2.35332300206892e-5 EC000430
126.056 C03239 M-H2O-H[-] 17.833 2.35332300206892e-5 EC000430
126.056 C04076 M-H2O-H[-] 17.833 2.35332300206892e-5 EC000430
128.0716 C00408 M-H[-] 19.977 1.23533230009798e-4 EC000242
128.0716 C05936 M-H[-] 19.977 1.23533230009798e-4 EC000242
117.0193 C00042 M-H[-] 10.096 2.35332300064783e-5 EC00029
133.0142 C00042 M-H+O[-] 10.137 3.35332300096525e-5 EC00029
152.99589 C00042 M+Cl[-] 9.537 3.90000000010104e-4 EC00029
163.02440000000001 C00042 M+HCOO[-] 10.02 1.55000000006567e-4 EC00029
177.0401 C00042 M+CH3COO[-] 9.724 2.04999999994016e-4 EC00029
117.0193 C02170 M-H[-] 10.096 2.35332300064783e-5 EC00029
133.0142 C02170 M-H+O[-] 10.137 3.35332300096525e-5 EC00029
152.99589 C02170 M+Cl[-] 9.537 3.90000000010104e-4 EC00029
163.02440000000001 C02170 M+HCOO[-] 10.02 1.55000000006567e-4 EC00029
177.0401 C02170 M+CH3COO[-] 9.724 2.04999999994016e-4 EC00029
152.9957 C00042 M+Cl[-] 2.879 2.00000000006639e-4 EC00030
158.04582 C00042 M+ACN-H[-] 3.132 4.85332300002028e-5 EC00030
99.0088 C00042 M-H2O-H[-] 2.661 7.64667699826305e-5 EC00030
152.9957 C02170 M+Cl[-] 2.879 2.00000000006639e-4 EC00030
158.04582 C02170 M+ACN-H[-] 3.132 4.85332300002028e-5 EC00030
99.0088 C02170 M-H2O-H[-] 2.661 7.64667699826305e-5 EC00030
```

Summary

Performing metabolomics analyses with R allows you to:

- Work efficiently
- Process larger amounts of data in less time
- Carry out several different binary comparisons with ease

Some potential introductory courses to R on LinkedIn Learning

Learning R
https://www.linkedin.com/learning/learning-r-2/r-for-data-science?u=76177458

R essential training: Wrangling and Visualizing data
https://www.linkedin.com/learning/r-essential-training-wrangling-and-visualizing-data/make-your-data-make-sense?u=76177458

More in-depth on ggplot2 in R
https://www.linkedin.com/learning/data-visualization-in-r-with-ggplot2/welcome?u=76177458

More in-depth of the R tidyverse
https://www.linkedin.com/learning/learning-the-r-tidyverse/welcome?u=76177458

Workshop/Tutorial with R for metabolomics

Date TBC, likely February or March

Assumed basic knowledge of R – see tutorials on previous page

Help you get started and show you how to set up code that allows flexible analysis