# Metabolomic Data Analysis with MetaboAnalystR

Name: guest6238390083494894 50

February 7, 2018

## 1 Background

Understanding the functional importance of metabolites in untargeted metabolomics is limited due to challenges with metabolite identication. to reduce problems associated with compound misidentification and thereby pathway misinterpretation is to shift the unit of analysis from individual compounds to individual pathways. In particular, the mummichog algorithm (5) offers an elegant and efficient implementation of this concept. Mummichog bypasses the bottleneck of metabolite identification prior to pathway analysis by leveraging a priori pathway and network knowledge to directly infer biological activity based on MS peaks. Due of its popularity and repeated user requests, we have implemented the mummichog algorithm (version 1.0.10) from Li et al. 2013, which has been carefully translated from the Python programming language to R, and includes a expanded knowledgebase of 21 organisms for pathway analysis. In particular, this module by-passes the bottle-neck of metabolite identification prior to pathway analysis, leveraging a priori knowledge from genome-scale metabolic models and KEGG metabolic pathways. For instance, conventional approaches require statistically significant metabolites to be identified prior to pathway analysis, which can be incredibly time-consuming, whereas here, m/z features are used in conjuction with metabolic models/pathways to directly infer pathway level knowledge from a m/z features. For further details, please refer to Li et al. 2013 (PMC3701697).

## 2 Overview

The MS Peaks to Pathways module consists of three steps - uploading the user's data, selection of a pathway library, and pathway analysis.

## 3 Data Input

The MS Peaks to Pathways module accepts either a list of significant m/z features, a list of all m/z features, or a peak intensity table. The format of the data must be specified, identifying whether the samples are in rows or columns, and whether or not the data is paired. The data may either be .csv or .txt files.

#### 3.0.1 MS Peaks to Pathways: Reading Data

The data must be uploaded as a three column table containing the m/z features, p-values, and statistical scores (e.g. t-scores or fold-change values).

A total of 3934 m/z features were found in your uploaded data.

### 3.0.2 Parameters

Users also need to specify the mass accuracy, the ion mode (positive or negative), and the p-value cutoff to delineate between significantly enriched and non-significantly enriched m/z features. Currently, MetaboAnalyst 4.0 only supports the handling of peaks obtained from high-resolution MS instruments such as Orbitrap, or Fourier Transform (FT)-MS instruments as recommended by the original mummichog implementation.

The selected mass accuracy of your MS instrument in ppm is: ten; The selected mode of your MS instrument is: positive; The selected p-value cutoff is: 1e-04.

### 3.0.3 Library

The knowledge-base for this module consists of five genome-scale metabolic models obtained from the original Python implementation which have either been manually curated or downloaded from BioCyc, as well as an expanded library of 21 organisms derived from KEGG metabolic pathways. Users must select one of 21 KEGG pathway libraries, or one of five metabolic models.

The user's selected library is: $\text{hsa}_m fn$.

# 4 Output

The aim of of this module is to leverage the power of known metabolic models/pathways to gain functional insight directly from m/z features. There are three steps in this module, 1) Permutations: A list of metabolites (the same length as the number of significant m/z features) are inferred from the user's uploaded set of m/z features, considering all potential matches (isotopes/adducts). These tentative compounds are then mapped onto known metabolic pathways for the selected organism. For each pathway, a fisher's exact or hypergeometric p-value is calculated. 2) Step 1 is repeated multiple times to calculate the null distribution of p-values for all pathways, and is modeled as a Gamma distribution. 3) Following this, the significant m/z features are used to calculate the p-values for each pathway (Step 1). These p-values are then adjusted for the permutations.

# 5 Pathway Analysis Results Table

The output of the MS Peaks to Pathways module consists of a table of results containing ranked pathways that are enriched in the user-uploaded data. The table includes the total number of hits, their raw p-values (Fisher's exact test or Hypergeometric), their EASE score, and the p-value modeled on user data using a Gamma distribution.

Table 1: Results of the mummichog pathway analysis

| | Pathway total | Hits.total | Hits.sig | EASE | FET | Gamma |
|---|---|---|---|---|---|---|
| Tryptophan metabolism | 94.00 | 64.00 | 21.00 | 0.01 | 0.00 | 0.00 |
| Ascorbate (Vitamin C) and Aldarate Metabolism | 29.00 | 18.00 | 9.00 | 0.01 | 0.00 | 0.00 |
| Aminosugars metabolism | 69.00 | 29.00 | 12.00 | 0.01 | 0.00 | 0.00 |
| Nitrogen metabolism | 6.00 | 4.00 | 4.00 | 0.02 | 0.00 | 0.01 |
| N-Glycan biosynthesis | 48.00 | 14.00 | 7.00 | 0.03 | 0.01 | 0.01 |
| Pyrimidine metabolism | 70.00 | 43.00 | 14.00 | 0.05 | 0.02 | 0.01 |
| Vitamin B3 (nicotinate and nicotinamide) metabolism | 28.00 | 19.00 | 8.00 | 0.05 | 0.02 | 0.01 |
| Sialic acid metabolism | 107.00 | 28.00 | 10.00 | 0.06 | 0.03 | 0.01 |
| Alanine and Aspartate Metabolism | 30.00 | 17.00 | 7.00 | 0.08 | 0.03 | 0.01 |
| Glutathione Metabolism | 19.00 | 10.00 | 5.00 | 0.10 | 0.03 | 0.01 |
| Hexose phosphorylation | 20.00 | 18.00 | 7.00 | 0.10 | 0.04 | 0.01 |
| Arginine and Proline Metabolism | 45.00 | 31.00 | 10.00 | 0.11 | 0.05 | 0.01 |
| Glycosphingolipid biosynthesis - ganglioseries | 62.00 | 11.00 | 5.00 | 0.13 | 0.04 | 0.01 |
| Glutamate metabolism | 15.00 | 11.00 | 5.00 | 0.13 | 0.04 | 0.01 |
| Methionine and cysteine metabolism | 94.00 | 46.00 | 13.00 | 0.14 | 0.08 | 0.01 |
| Purine metabolism | 80.00 | 51.00 | 14.00 | 0.14 | 0.08 | 0.01 |
| Parathio degradation | 6.00 | 4.00 | 3.00 | 0.16 | 0.02 | 0.01 |
| Vitamin B9 (folate) metabolism | 33.00 | 12.00 | 5.00 | 0.17 | 0.06 | 0.01 |
| Glycosphingolipid biosynthesis - globoseries | 16.00 | 8.00 | 4.00 | 0.18 | 0.05 | 0.01 |
| Starch and Sucrose Metabolism | 33.00 | 15.00 | 5.00 | 0.31 | 0.13 | 0.01 |
| N-Glycan Degradation | 16.00 | 6.00 | 3.00 | 0.31 | 0.08 | 0.02 |
| Linoleate metabolism | 46.00 | 21.00 | 6.00 | 0.36 | 0.19 | 0.02 |
| Urea cycle/amino group metabolism | 85.00 | 42.00 | 10.00 | 0.39 | 0.26 | 0.02 |
| Glycosphingolipid metabolism | 67.00 | 28.00 | 7.00 | 0.43 | 0.27 | 0.02 |
| Aspartate and asparagine metabolism | 114.00 | 61.00 | 13.00 | 0.48 | 0.36 | 0.03 |
| Caffeine metabolism | 11.00 | 9.00 | 3.00 | 0.53 | 0.23 | 0.04 |
| Lipoate metabolism | 8.00 | 4.00 | 2.00 | 0.57 | 0.16 | 0.04 |
| Vitamin B1 (thiamin) metabolism | 20.00 | 10.00 | 3.00 | 0.59 | 0.29 | 0.05 |
| Drug metabolism - cytochrome P450 | 53.00 | 38.00 | 8.00 | 0.59 | 0.43 | 0.05 |
| TCA cycle | 31.00 | 16.00 | 4.00 | 0.60 | 0.36 | 0.05 |
| Pentose phosphate pathway | 37.00 | 33.00 | 7.00 | 0.61 | 0.43 | 0.05 |
| Porphyrin metabolism | 43.00 | 22.00 | 5.00 | 0.62 | 0.40 | 0.05 |
| Alkaloid biosynthesis II | 10.00 | 5.00 | 2.00 | 0.65 | 0.24 | 0.06 |
| Glycosphingolipid biosynthesis - lactoseries | 14.00 | 5.00 | 2.00 | 0.65 | 0.24 | 0.06 |
| O-Glycan biosynthesis | 16.00 | 5.00 | 2.00 | 0.65 | 0.24 | 0.06 |
| Tyrosine metabolism | 160.00 | 87.00 | 16.00 | 0.69 | 0.59 | 0.07 |
| Glycolysis and Gluconeogenesis | 49.00 | 36.00 | 7.00 | 0.70 | 0.53 | 0.08 |
| C5-Branched dibasic acid metabolism | 10.00 | 6.00 | 2.00 | 0.71 | 0.32 | 0.08 |
| Chondroitin sulfate degradation | 37.00 | 6.00 | 2.00 | 0.71 | 0.32 | 0.08 |
| Blood Group Biosynthesis | 44.00 | 6.00 | 2.00 | 0.71 | 0.32 | 0.08 |
| Glycosphingolipid biosynthesis - neolactoseries | 16.00 | 6.00 | 2.00 | 0.71 | 0.32 | 0.08 |
| Keratan sulfate biosynthesis | 66.00 | 6.00 | 2.00 | 0.71 | 0.32 | 0.08 |
| Glycerophospholipid metabolism | 156.00 | 37.00 | 7.00 | 0.72 | 0.56 | 0.09 |
| Pentose and Glucuronate Interconversions | 15.00 | 13.00 | 3.00 | 0.73 | 0.46 | 0.09 |
| Heparan sulfate degradation | 34.00 | 7.00 | 2.00 | 0.77 | 0.39 | 0.11 |
| Lysine metabolism | 52.00 | 28.00 | 5.00 | 0.80 | 0.63 | 0.13 |
| Keratan sulfate degradation | 68.00 | 8.00 | 2.00 | 0.81 | 0.46 | 0.13 |
| Vitamin B5 - CoA biosynthesis from pantothenate | 12.00 | 8.00 | 2.00 | 0.81 | 0.46 | 0.13 |
| Carbon fixation | 10.00 | 8.00 | 2.00 | 0.81 | 0.46 | 0.13 |
| Vitamin E metabolism | 54.00 | 16.00 | 3.00 | 0.83 | 0.61 | 0.15 |
| Drug metabolism - other enzymes | 31.00 | 16.00 | 3.00 | 0.83 | 0.61 | 0.15 |
| Histidine metabolism | 33.00 | 19.00 | 3.00 | 0.90 | 0.73 | 0.22 |
| Butanoate metabolism | 34.00 | 27.00 | 4.00 | 0.91 | 0.78 | 0.24 |
| Xenobiotics metabolism | 110.00 | 70.00 | 10.00 | 0.93 | 0.88 | 0.29 |
| Squalene and cholesterol biosynthesis | 55.00 | 22.00 | 3.00 | 0.94 | 0.81 | 0.30 |
| Selenoamino acid metabolism | 35.00 | 22.00 | 3.00 | 0.94 | 0.81 | 0.30 |
| Glycine, serine, alanine and threonine metabolism | 88.00 | 51.00 | 7.00 | 0.94 | 0.88 | 0.30 |
| Galactose metabolism | 41.00 | 30.00 | 4.00 | 0.94 | 0.85 | 0.31 |
| Biopterin metabolism | 22.00 | 14.00 | 2.00 | 0.95 | 0.77 | 0.32 |
| Saturated fatty acids beta-oxidation | 36.00 | 14.00 | 2.00 | 0.95 | 0.77 | 0.32 |
| Beta-Alanine metabolism | 20.00 | 14.00 | 2.00 | 0.95 | 0.77 | 0.32 |
| Putative anti-Inflammatory metabolites formation from EPA | 27.00 | 15.00 | 2.00 | 0.96 | 0.81 | 0.35 |
| Phosphatidylinositol phosphate metabolism | 59.00 | 24.00 | 3.00 | 0.96 | 0.86 | 0.35 |
| Carnitine shuttle | 72.00 | 16.00 | 2.00 | 0.96 | 0.84 | 0.38 |
| Propanoate metabolism | 31.00 | 20.00 | 2.00 | 0.98 | 0.92 | 0.50 |
| Leukotriene metabolism | 92.00 | 30.00 | 3.00 | 0.99 | 0.94 | 0.50 |
| Prostaglandin formation from arachidonate | 78.00 | 42.00 | 4.00 | 0.99 | 0.97 | 0.56 |
| Androgen and estrogen biosynthesis and metabolism | 95.00 | 53.00 | 5.00 | 0.99 | 0.98 | 0.61 |
| Fructose and mannose metabolism | 33.00 | 26.00 | 2.00 | 1.00 | 0.97 | 0.63 |
| Valine, leucine and isoleucine degradation | 65.00 | 40.00 | 3.00 | 1.00 | 0.99 | 0.69 |
| Arachidonic acid metabolism | 95.00 | 44.00 | 3.00 | 1.00 | 0.99 | 0.74 |
| C21-steroid hormone biosynthesis and metabolism | 112.00 | 57.00 | 2.00 | 1.00 | 1.00 | 0.93 |
| Hyaluronan Metabolism | 8.00 | 4.00 | 1.00 | 1.00 | 0.57 | 1.00 |
| Pyruvate Metabolism | 20.00 | 12.00 | 1.00 | 1.00 | 0.92 | 1.00 |
| Vitamin H (biotin) metabolism | 5.00 | 4.00 | 1.00 | 1.00 | 0.57 | 1.00 |
| Glyoxylate and Dicarboxylate Metabolism | 12.00 | 6.00 | 1.00 | 1.00 | 0.72 | 1.00 |
| De novo fatty acid biosynthesis | 106.00 | 19.00 | 1.00 | 1.00 | 0.98 | 1.00 |
| Fatty Acid Metabolism | 63.00 | 10.00 | 1.00 | 1.00 | 0.88 | 1.00 |
| CoA Catabolism | 7.00 | 5.00 | 1.00 | 1.00 | 0.65 | 1.00 |
| D4&E4-neuroprostanes formation | 37.00 | 19.00 | 1.00 | 1.00 | 0.98 | 1.00 |
| Dynorphin metabolism | 8.00 | 4.00 | 1.00 | 1.00 | 0.57 | 1.00 |
| Vitamin B6 (pyridoxine) metabolism | 11.00 | 7.00 | 1.00 | 1.00 | 0.77 | 1.00 |

# 6 MS Peaks to Pathway Output: Compound Matching Table

The output of the MS Peaks to Pathways module also consists of a comprehensive table containing the compound matching information for all user-uploaded m/z features. The table has four columns, containing the Query.Mass of each feature, the predicted Matched.Compound for each feature, the Matched.Form, and the Mass.Diff. As the file can be very long (>40 pages), please download it separately on the Downloads page of MetaboAnalyst.

# 7 Appendix: R Command History

```
[1] "InitDataObjects(\"mass_all\", \"mummichog\", FALSE)"
[2] "mSet<-Read.PeakListData(mSet, \"Replacing_with_your_file_path\");"
[3] "mSet<-UpdateMummichogParameters(mSet, \"ten\", \"positive\", 1.0E-4);"
[4] "mSet<-SanityCheckMummichogData(mSet)"
[5] "mSet<-SetMass.PathLib(mSet, \"hsa_mfn\")"
[6] "mSet<-PerformMummichog(mSet, \"fisher\", \"gamma\")"
[7] "mSet<-SaveTransformedData(mSet)"
```

---

The report was generated on Wed Feb 7 16:12:35 2018 with R version 3.4.1 (2017-06-30).