Dividing up the data

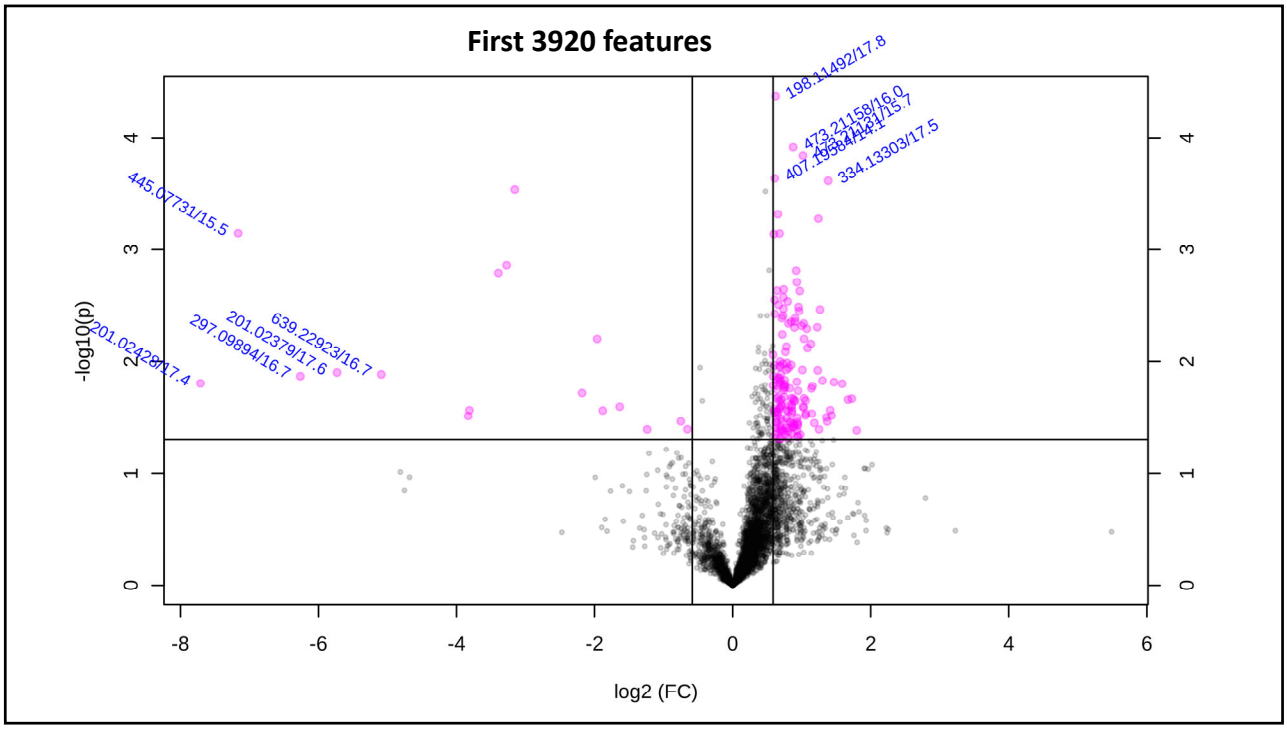
Stephen Barnes

1

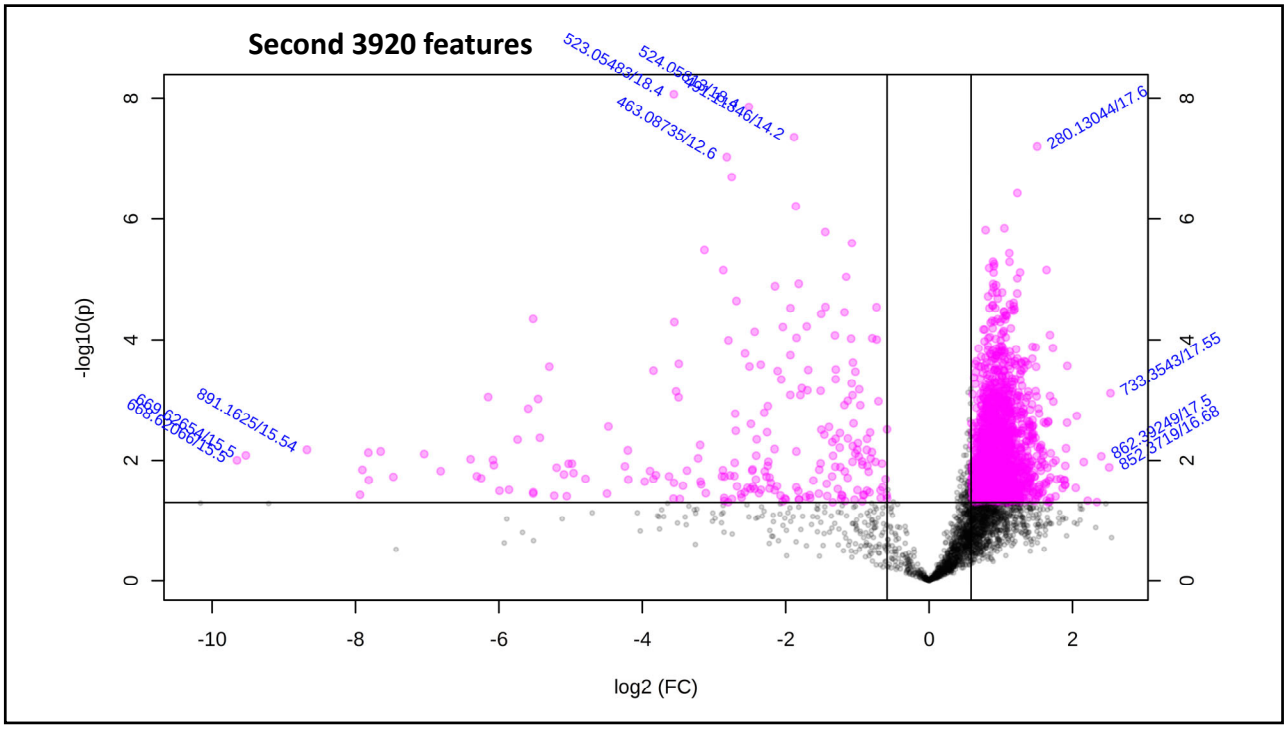
Analysis of the control-genistein data

- **The Excel download from the XCMS analysis contained 7840 ion features**
  - Too many to load them onto MetaboAnalyst (restricted to 5000 features)
- **The data on the 5<RT<25 min sheet was ordered (highest to lowest) according to the values in the control mean column**
  - These ordered data were divided into two sets – lines 1-3921 (the high group) and the line 3922-7841 (the low group) – on new sheets.
  - Added line 1 from the high group to the top of the 3922-7841 sheet (this identifies the m/z, RT, and sample names)
  - For each new sheet, .csv files were created for each sample, add to the control and genistein folders, zipped up and analyzed on Metaboanalyst.
  - Note, the ion selection is now NONE instead of INTERQUARTILE RANGE

2

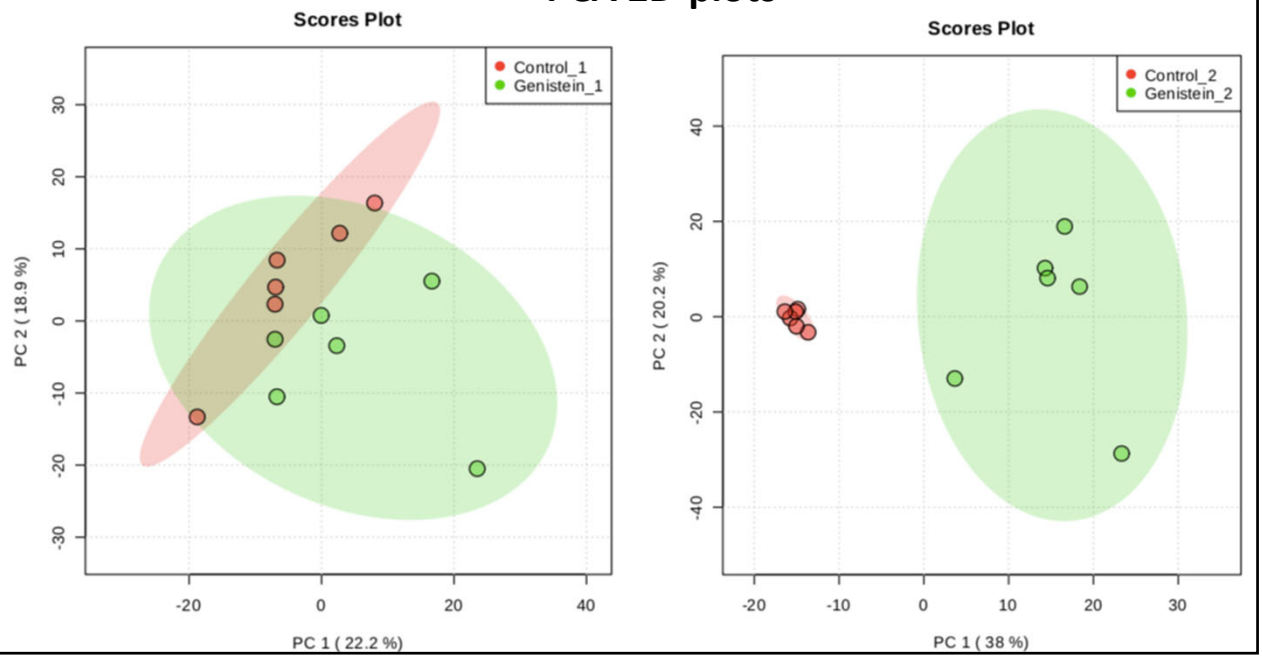

First 3920 features
198.11492/17.8
473.21158/16.0
334.13303/17.5
445.07731/15.5
639.22923/16.7
201.02379/17.6
297.09894/16.7
-log10(p)
log2 (FC)

3


Second 3920 features
523.05483/18.4
463.08735/12.6
280.13044/17.6
733.3543/17.55
862.39249/17.5
852.3719/16.68
891.1625/15.54
-log10(p)
log2 (FC)

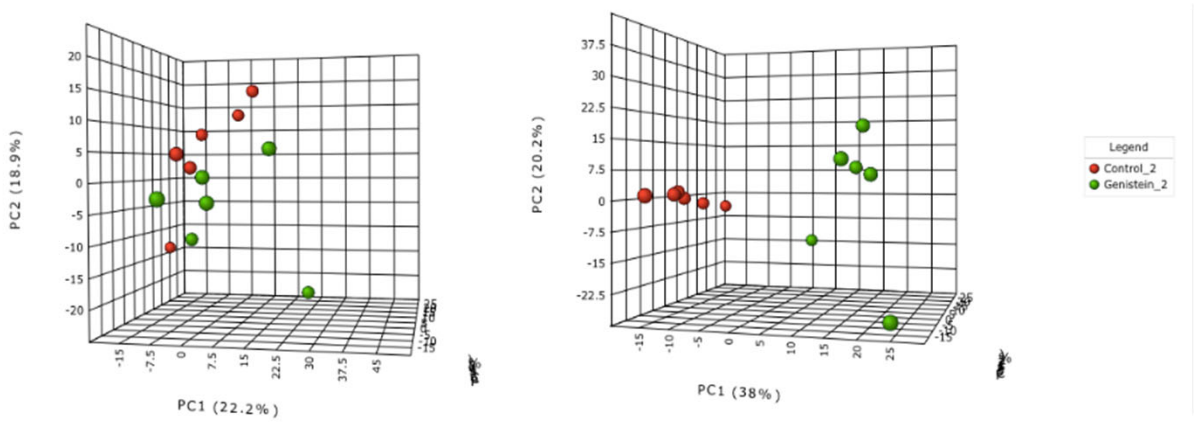
4

PCA 2D-plots



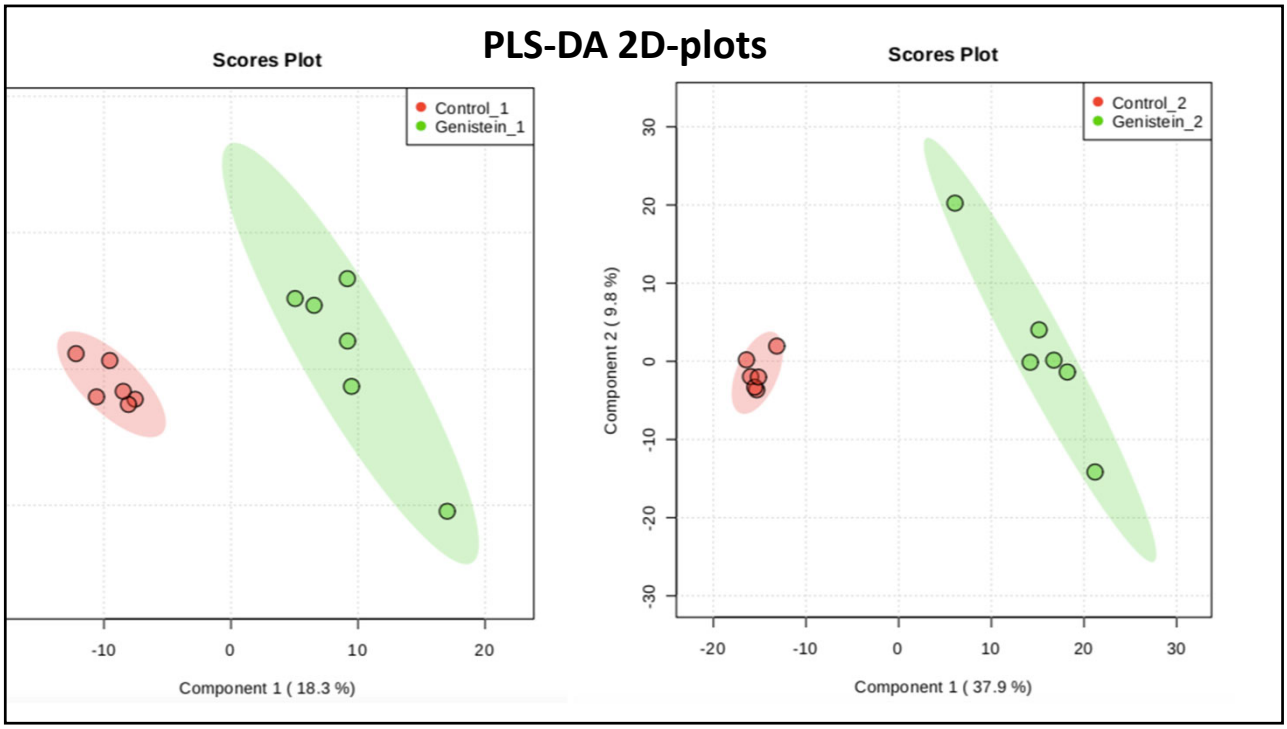
5

PCA 3D-plots



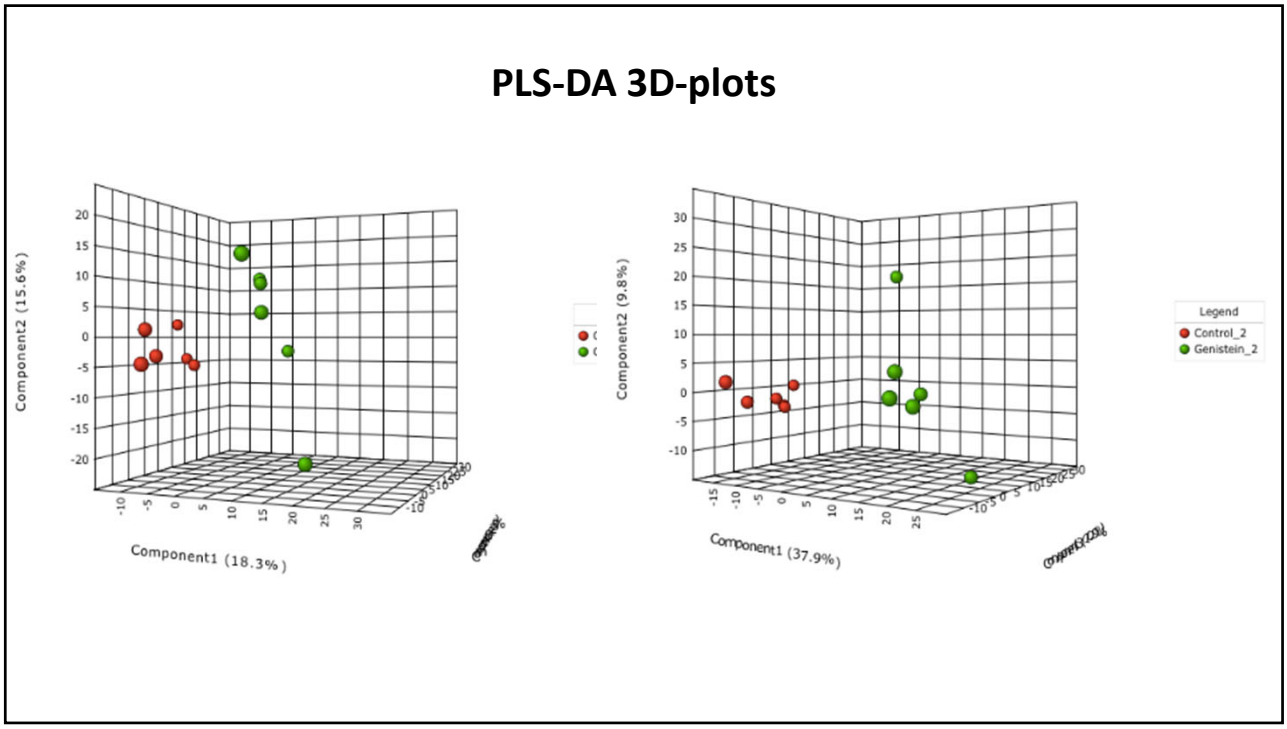
6

PLS-DA 2D-plots

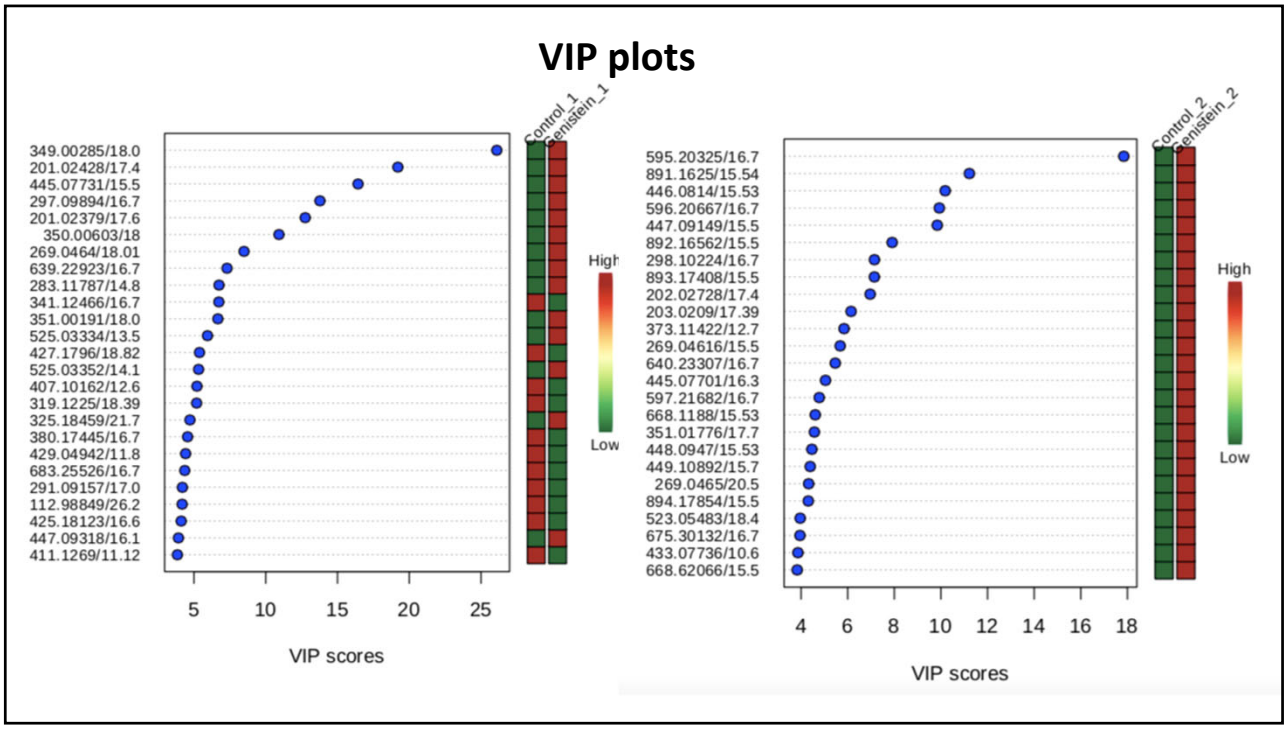


7

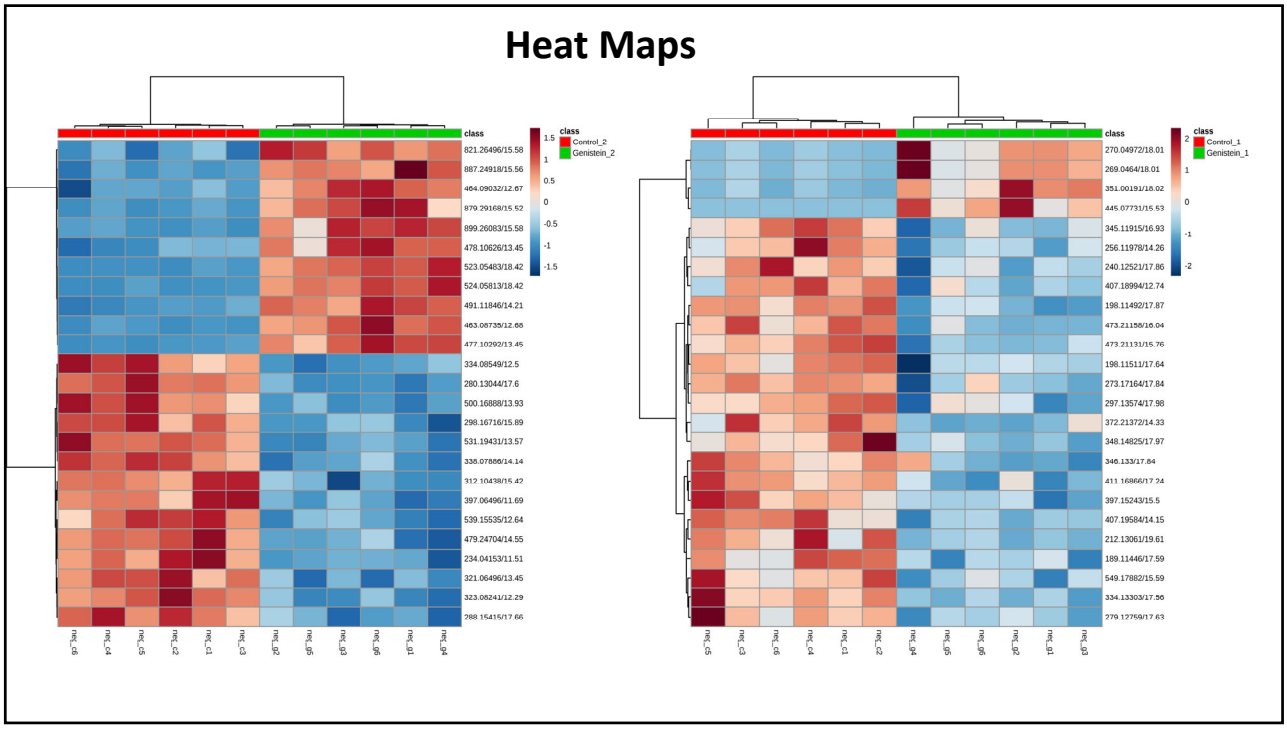
PLS-DA 3D-plots



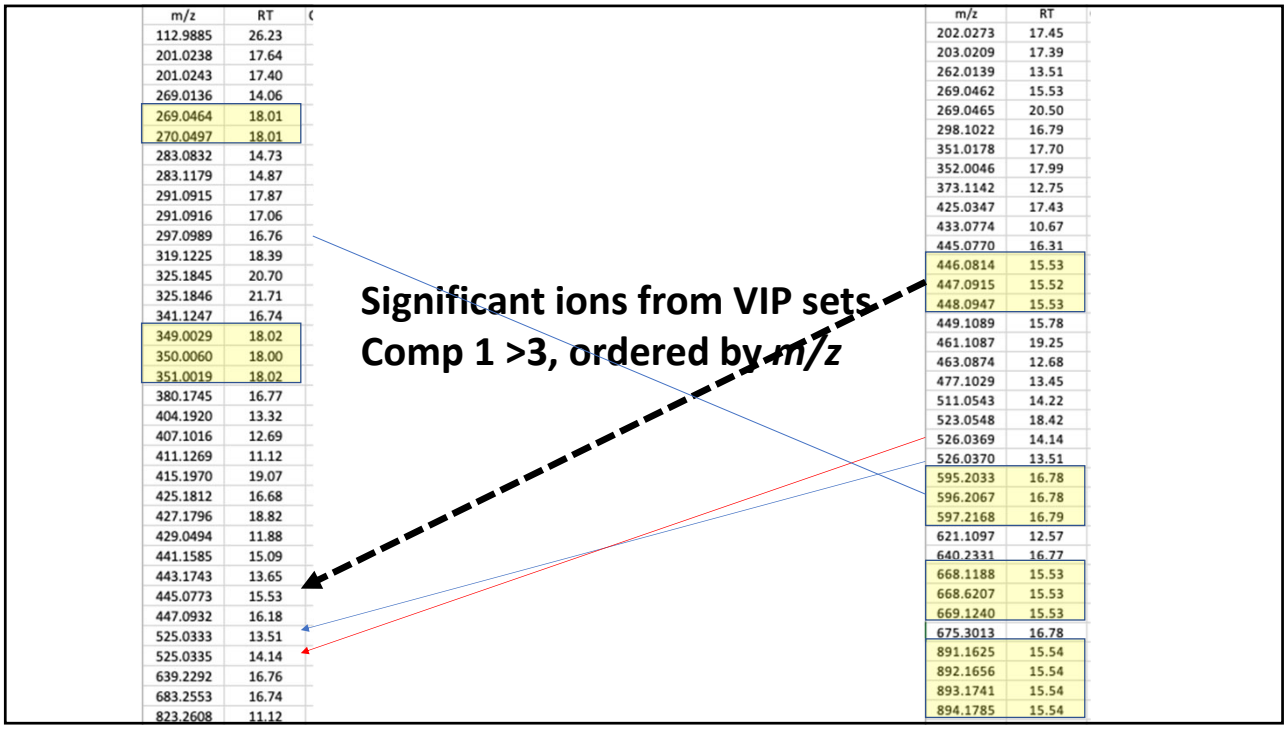
8


VIP plots
349.00285/18.0
201.02428/17.4
445.07731/15.5
297.09894/16.7
201.02379/17.6
350.00603/18
269.0464/18.01
639.22923/16.7
283.11787/14.8
341.12466/16.7
351.00191/18.0
525.03334/13.5
427.1796/18.82
525.03352/14.1
407.10162/12.6
319.1225/18.39
325.18459/21.7
380.17445/16.7
429.04942/11.8
683.25526/16.7
291.09157/17.0
112.98849/26.2
425.18123/16.6
447.09318/16.1
411.1269/11.12
Control_1
Genistein_1
High
Low
5
10
15
20
25
VIP scores
595.20325/16.7
891.1625/15.54
446.0814/15.53
596.20667/16.7
447.09149/15.5
892.16562/15.5
298.10224/16.7
893.17408/15.5
202.02728/17.4
203.0209/17.39
373.11422/12.7
269.04616/15.5
640.23307/16.7
445.07701/16.3
597.21682/16.7
668.1188/15.53
351.01776/17.7
448.0947/15.53
449.10892/15.7
269.0465/20.5
894.17854/15.5
523.05483/18.4
675.30132/16.7
433.07736/10.6
668.62066/15.5
Control_2
Genistein_2
High
Low
4
6
8
10
12
14
16
18
VIP scores

9


Heat Maps
class
Control_2
Genistein_2
821.26496/15.58
887.24918/15.56
464.09032/12.67
879.29168/15.52
899.26083/15.58
478.10626/13.45
523.05483/18.42
524.05813/18.42
491.11846/14.21
463.08735/12.66
477.10292/13.45
334.08549/12.5
280.13044/17.6
500.16888/13.93
298.16716/15.89
531.19431/13.57
336.07886/14.14
397.06496/11.69
539.15535/12.64
479.24704/14.55
234.04153/11.51
321.06496/13.45
323.08241/12.29
class
Control_1
Genistein_1
270.04972/18.01
269.0464/18.01
351.00191/18.02
446.07731/15.53
345.11915/16.93
256.11978/14.26
240.12521/17.86
407.18994/12.74
198.11492/17.87
273.17164/17.84
297.13574/17.98
372.21372/14.33
348.14825/17.97
346.133/17.84
397.15243/15.5
407.19584/14.15
212.13061/19.61
189.11446/17.59
549.17882/15.59

10



m/z	RT
112.9885	26.23
201.0238	17.64
201.0243	17.40
269.0136	14.06
269.0464	18.01
270.0497	18.01
283.0832	14.73
283.1179	14.87
291.0915	17.87
291.0916	17.06
297.0989	16.76
319.1225	18.39
325.1845	20.70
325.1846	21.71
341.1247	16.74
349.0029	18.02
350.0060	18.00
351.0019	18.02
380.1745	16.77
404.1920	13.32
407.1016	12.69
411.1269	11.12
415.1970	19.07
425.1812	16.68
427.1796	18.82
429.0494	11.88
441.1585	15.09
443.1743	13.65
445.0773	15.53
447.0932	16.18
525.0333	13.51
525.0335	14.14
639.2292	16.76
683.2553	16.74
823.2608	11.12

m/z	RT
202.0273	17.45
203.0209	17.39
262.0139	13.51
269.0462	15.53
269.0465	20.50
298.1022	16.79
351.0178	17.70
352.0046	17.99
373.1142	12.75
425.0347	17.43
433.0774	10.67
445.0770	16.31
446.0814	15.53
447.0915	15.52
448.0947	15.53
449.1089	15.78
461.1087	19.25
463.0874	12.68
477.1029	13.45
511.0543	14.22
523.0548	18.42
526.0369	14.14
526.0370	13.51
595.2033	16.78
596.2067	16.78
597.2168	16.79
621.1097	12.57
640.2331	16.77
668.1188	15.53
668.6207	15.53
669.1240	15.53
675.3013	16.78
891.1625	15.54
892.1656	15.54
893.1741	15.54
894.1785	15.54

11

Resolving the *m/z* 445, 668 and 891 ions

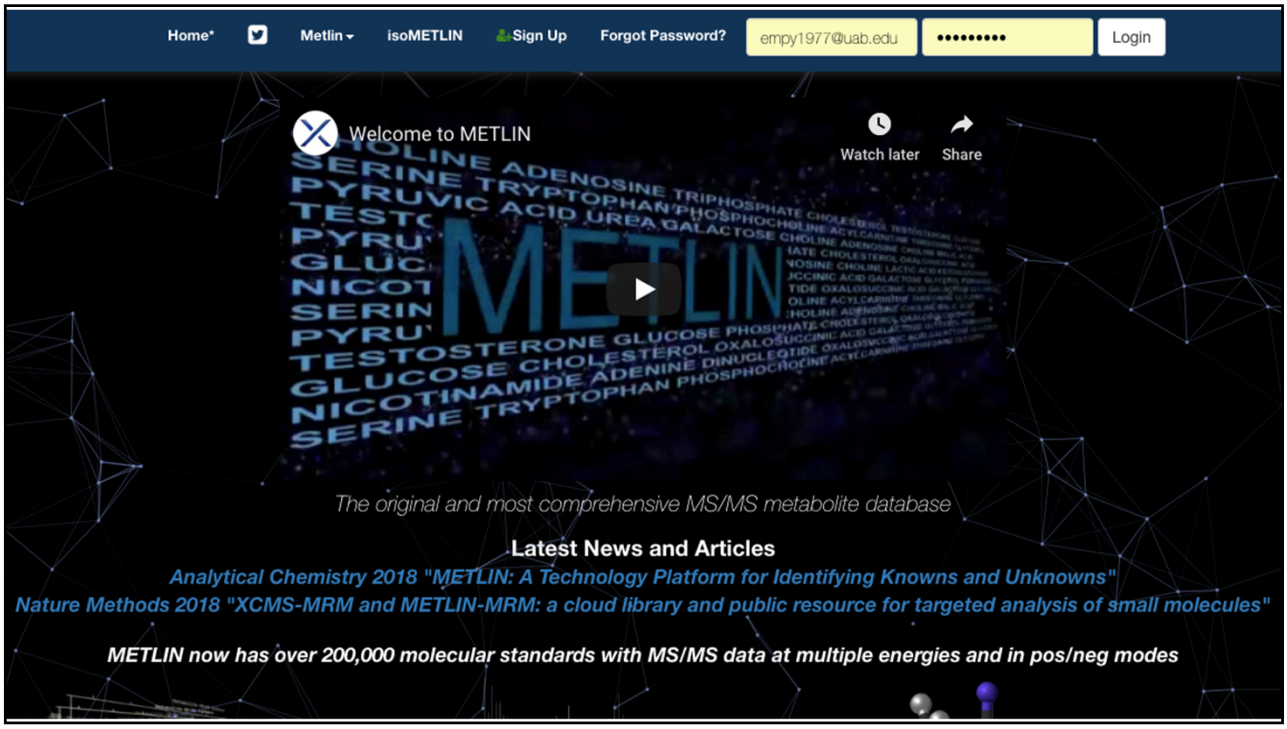
- ***m/z* 445.077 is genistein glucuronide $[M-H]^-$**
  - M is 446.086
- **$[2M-H]^-$ is *m/z* 892.172-1.007 = 891.165 (obs. *m/z* 891.163)**
  - The *m/z* 891-893 series are the 2M-H series with ^{13}C-isotopes
- **The *m/z* 668-669 series are doubly charged**
  - The ^{13}C-isotope spacing is 0.5
  - *m/z* 668.119 is $[m-2H]^{2-}$
  - m is therefore 2 x 668.119 + 2 x 1.0072 = 1336.238 + 2.014 = 1338.252
  - Divide 1338.252 by 446.086 = 3
- ***m/z* 668.119 is therefore $[3M-2H]^{2-}$**
- **All these ions are from genistein glucuronide**

12

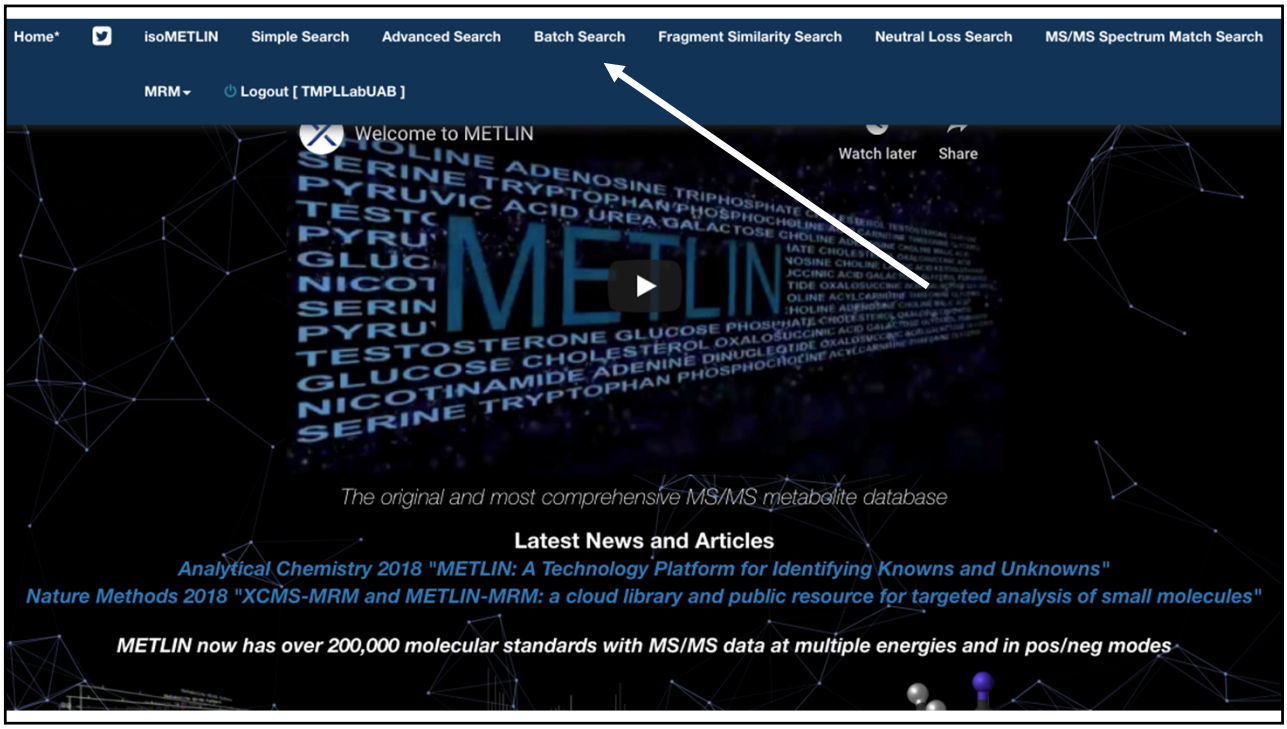
Now let's go to METLIN

Put METLIN in your browser
use your account information

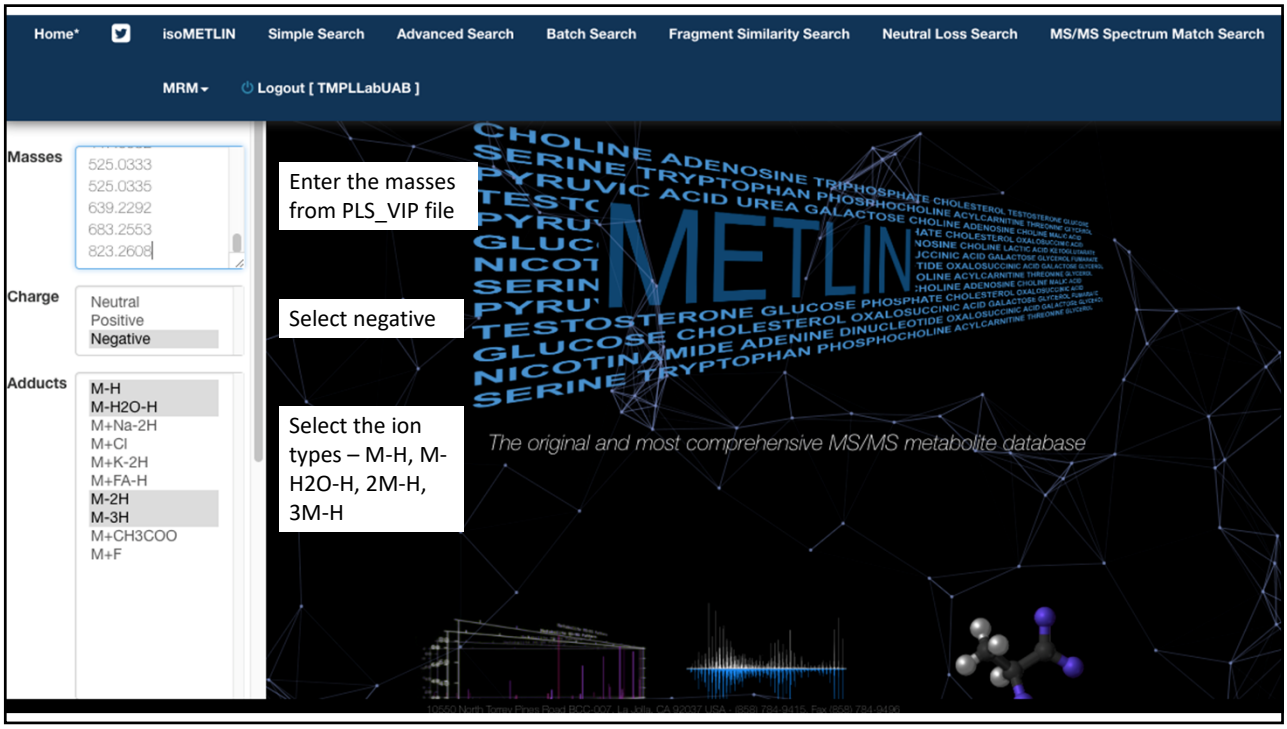
13



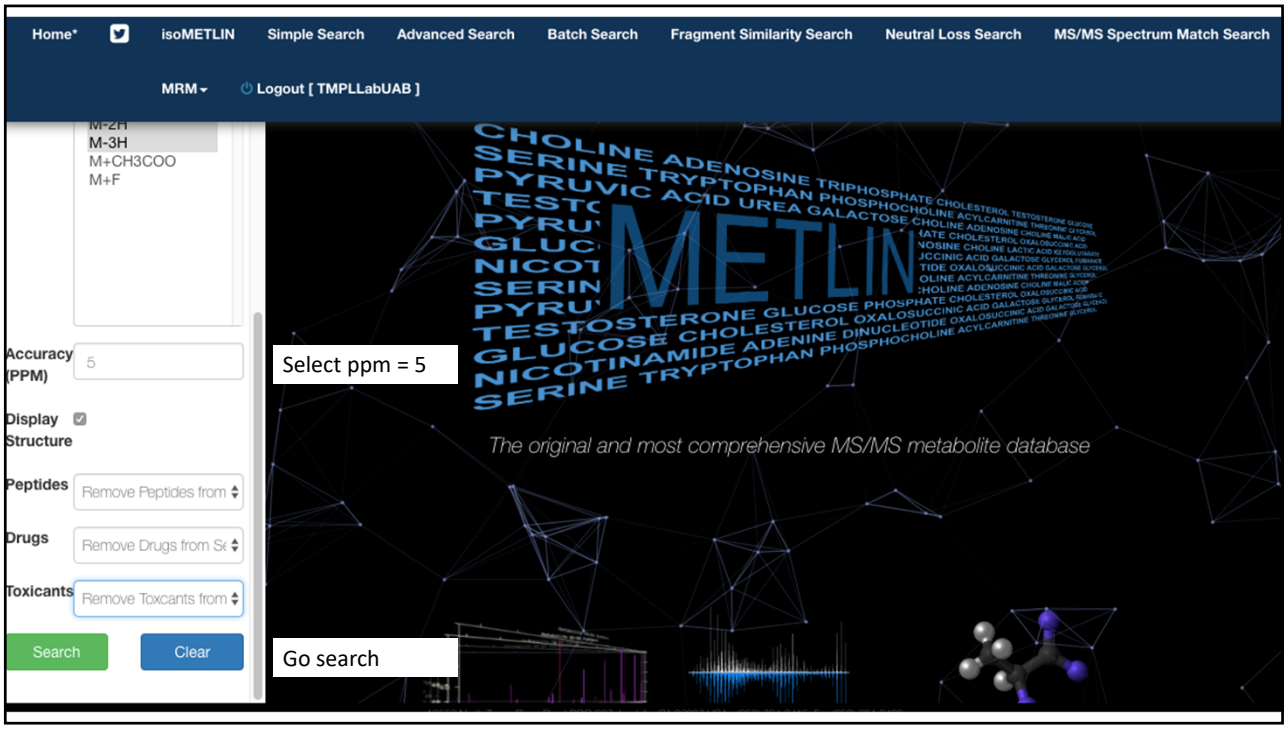
14


Home*
isoMETLIN
Simple Search
Advanced Search
Batch Search
Fragment Similarity Search
Neutral Loss Search
MS/MS Spectrum Match Search
MRM
Logout [TMPLLabUAB]
Welcome to METLIN
Watch later
Share
METLIN
The original and most comprehensive MS/MS metabolite database
Latest News and Articles
Analytical Chemistry 2018 "METLIN: A Technology Platform for Identifying Knowns and Unknowns"
Nature Methods 2018 "XCMS-MRM and METLIN-MRM: a cloud library and public resource for targeted analysis of small molecules"
METLIN now has over 200,000 molecular standards with MS/MS data at multiple energies and in pos/neg modes

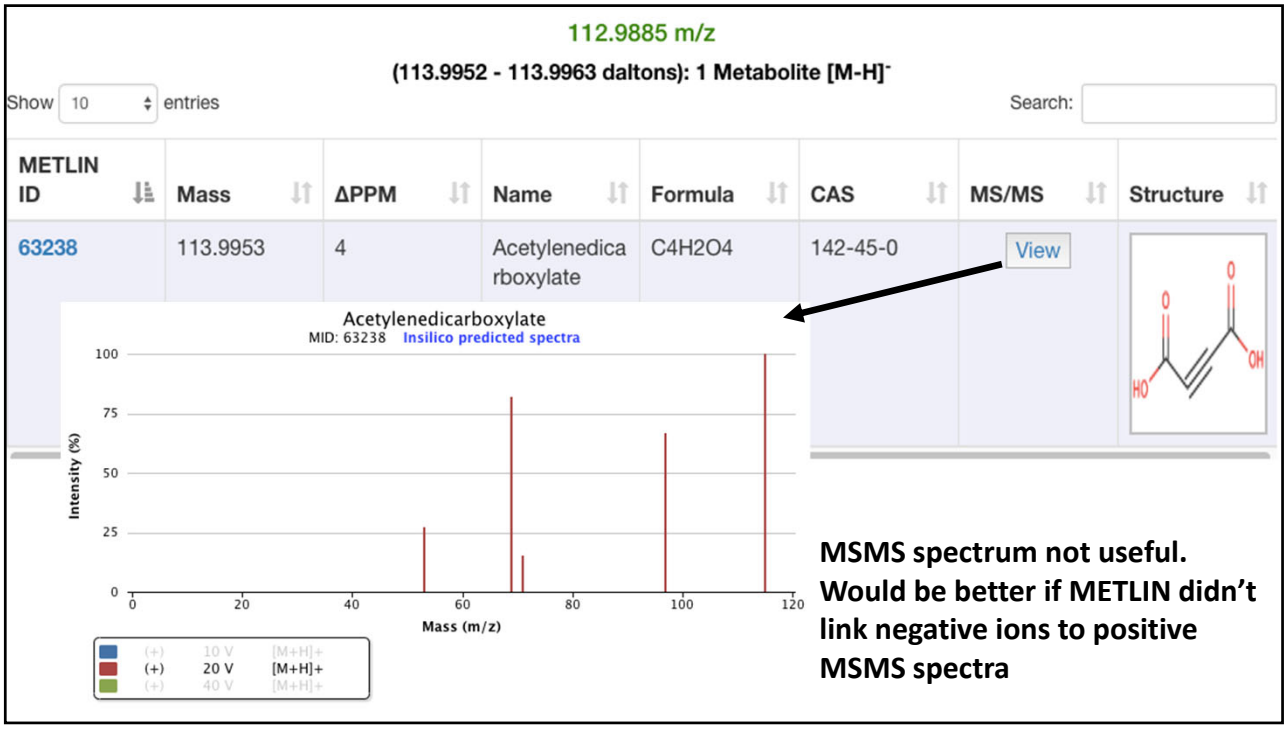
15


Home*
isoMETLIN
Simple Search
Advanced Search
Batch Search
Fragment Similarity Search
Neutral Loss Search
MS/MS Spectrum Match Search
MRM
Logout [TMPLLabUAB]
Masses
525.0333
525.0335
639.2292
683.2553
823.2608
Charge
Neutral
Positive
Negative
Adducts
M-H
M-H2O-H
M+Na-2H
M+Cl
M+K-2H
M+FA-H
M-2H
M-3H
M+CH3COO
M+F
Enter the masses from PLS_VIP file
Select negative
Select the ion types – M-H, M-H2O-H, 2M-H, 3M-H
METLIN
The original and most comprehensive MS/MS metabolite database

16


Home*
isoMETLIN
Simple Search
Advanced Search
Batch Search
Fragment Similarity Search
Neutral Loss Search
MS/MS Spectrum Match Search
MRM
Logout [TMPLLabUAB]
M-3H
M+CH3COO
M+F
Accuracy (PPM)
5
Display Structure
Peptides
Remove Peptides from
Drugs
Remove Drugs from Se
Toxicants
Remove Toxcants from
Search
Clear
METLIN
The original and most comprehensive MS/MS metabolite database
Select ppm = 5
Go search

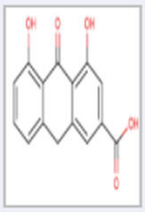
17


112.9885 m/z
(113.9952 - 113.9963 daltons): 1 Metabolite [M-H]-
Show 10 entries
Search:
METLIN ID
Mass
ΔPPM
Name
Formula
CAS
MS/MS
Structure
63238
113.9953
4
Acetylenedicarboxylate
C4H2O4
142-45-0
View
Acetylenedicarboxylate
MID: 63238 Insilico predicted spectra
Intensity (%)
Mass (m/z)
(+) 20 V [M+H]+
MSMS spectrum not useful. Would be better if METLIN didn't link negative ions to positive MSMS spectra

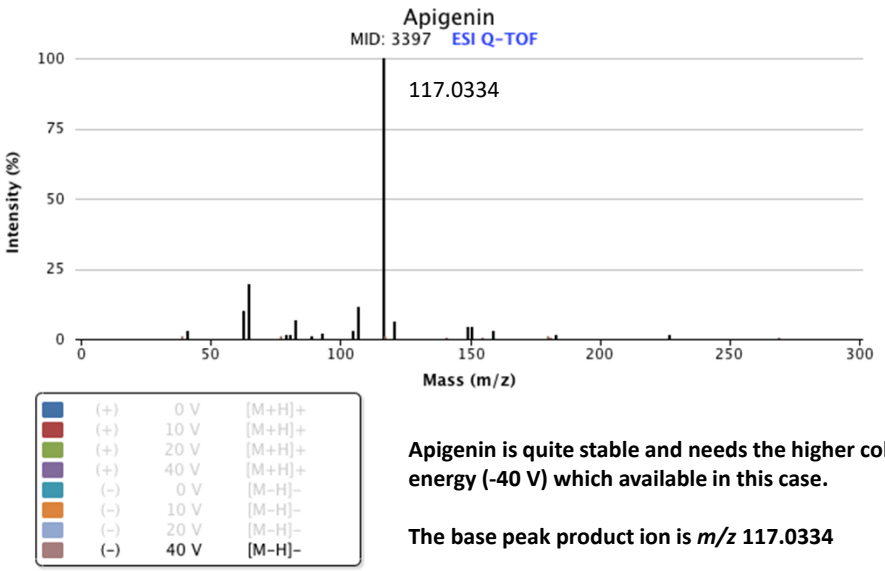
18

(270.0523 - 270.055 daltons): 29 Metabolites [M-H]⁻

Show 10 entries Search:

METLIN ID	Mass	ΔPPM	Name	Formula	CAS	MS/MS	Structure
2412	270.0528	3	Rhein-9-anthrone	C15H10O5	480-09-1	View	
3397	270.0528	3	Apigenin	C15H10O5	520-36-5	View	

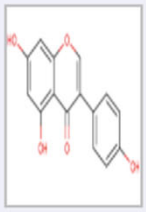
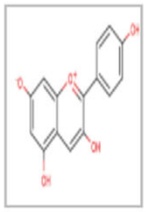
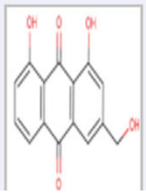
19



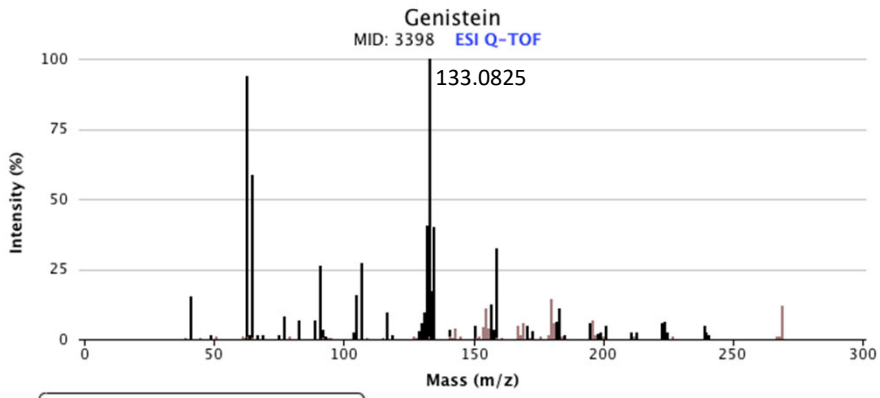
Apigenin is quite stable and needs the higher collision energy (-40 V) which available in this case.

The base peak product ion is *m/z* 117.0334

20

3398	270.0528	3	Genistein	C15H10O5	446-72-0	View	
3399	270.0528	3	Pelargonidin	C15H10O5	7690-51-9	View	
41039	270.0528	3	Aloe-emodin	C15H10O5	NA	View	

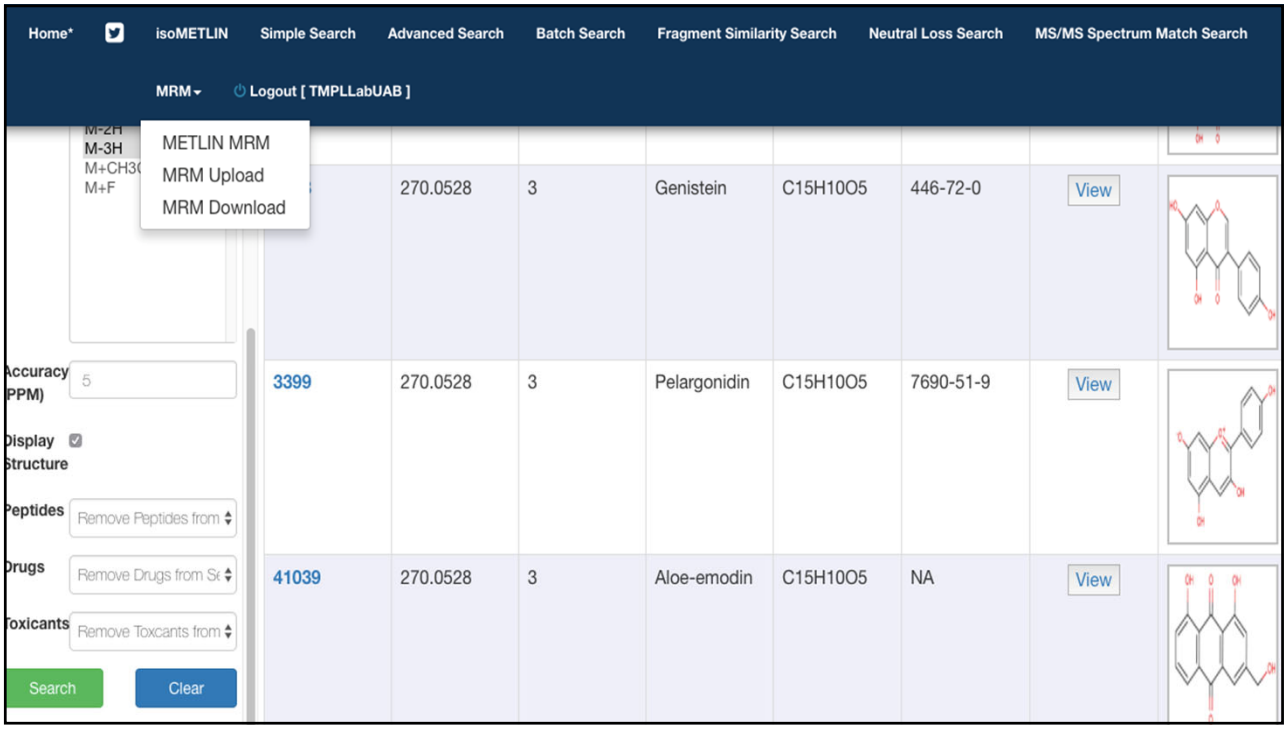
21



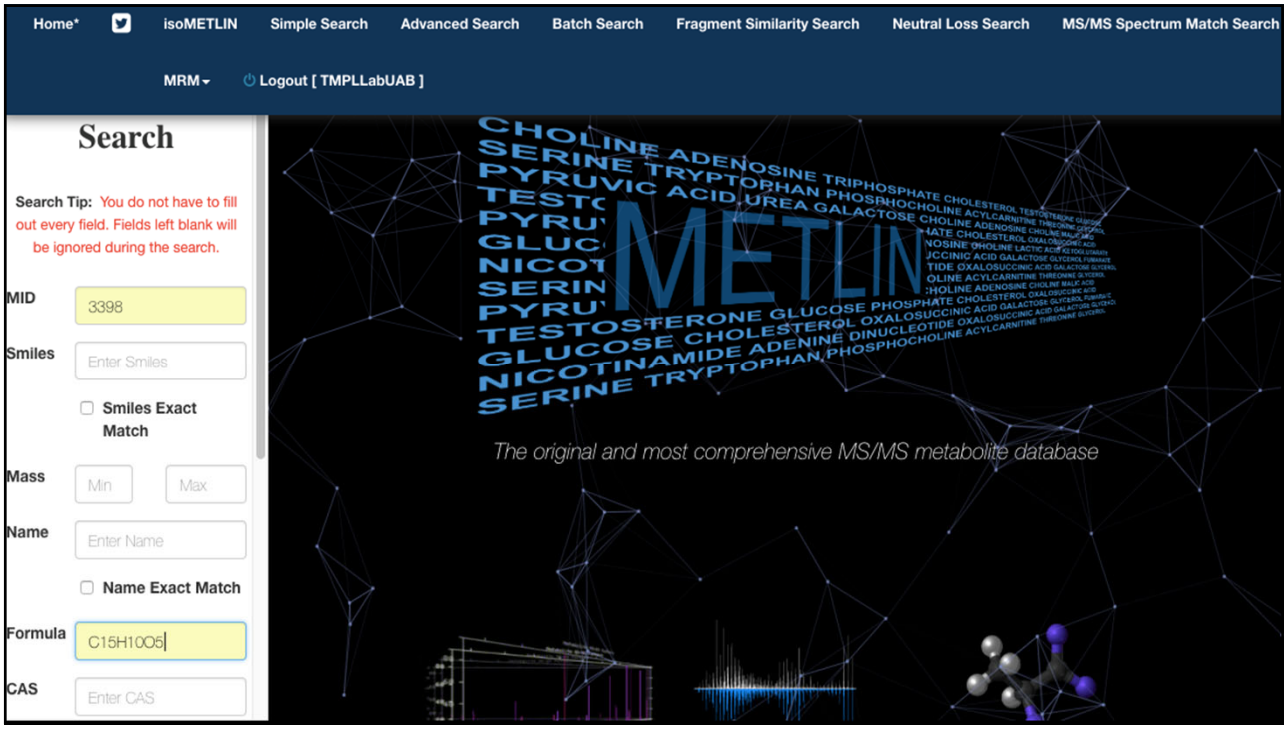
Genistein is quite stable and needs the higher collision energy (-40 V) which available in this case.

Its base peak product ion (*m/z* 133.0825) is different from its isomer apigenin (*m/z* 117.0334).

22

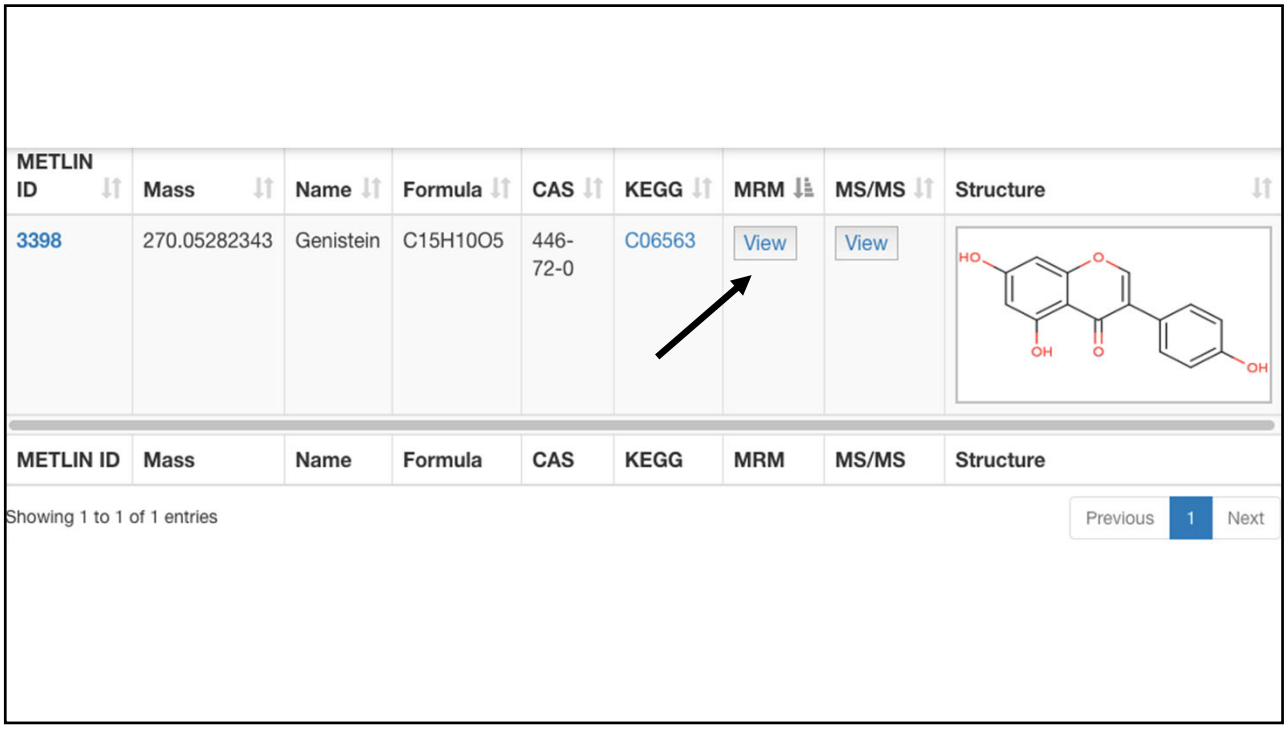

Home*
isoMETLIN
Simple Search
Advanced Search
Batch Search
Fragment Similarity Search
Neutral Loss Search
MS/MS Spectrum Match Search
MRM
Logout [TMPLLabUAB]
METLIN MRM
MRM Upload
MRM Download
M-3H
M+F
Accuracy (PPM)
5
Display Structure
Peptides
Remove Peptides from
Drugs
Remove Drugs from
Toxicants
Remove Toxcants from
Search
Clear
270.0528 3 Genistein C15H10O5 446-72-0 View
3399 270.0528 3 Pelargonidin C15H10O5 7690-51-9 View
41039 270.0528 3 Aloe-emodin C15H10O5 NA View

23


Home*
isoMETLIN
Simple Search
Advanced Search
Batch Search
Fragment Similarity Search
Neutral Loss Search
MS/MS Spectrum Match Search
MRM
Logout [TMPLLabUAB]
Search
Search Tip: You do not have to fill out every field. Fields left blank will be ignored during the search.
MID
3398
Smiles
Enter Smiles
Smiles Exact Match
Mass
Min
Max
Name
Enter Name
Name Exact Match
Formula
C15H10O5
CAS
Enter CAS
METLIN
The original and most comprehensive MS/MS metabolite database

24

25



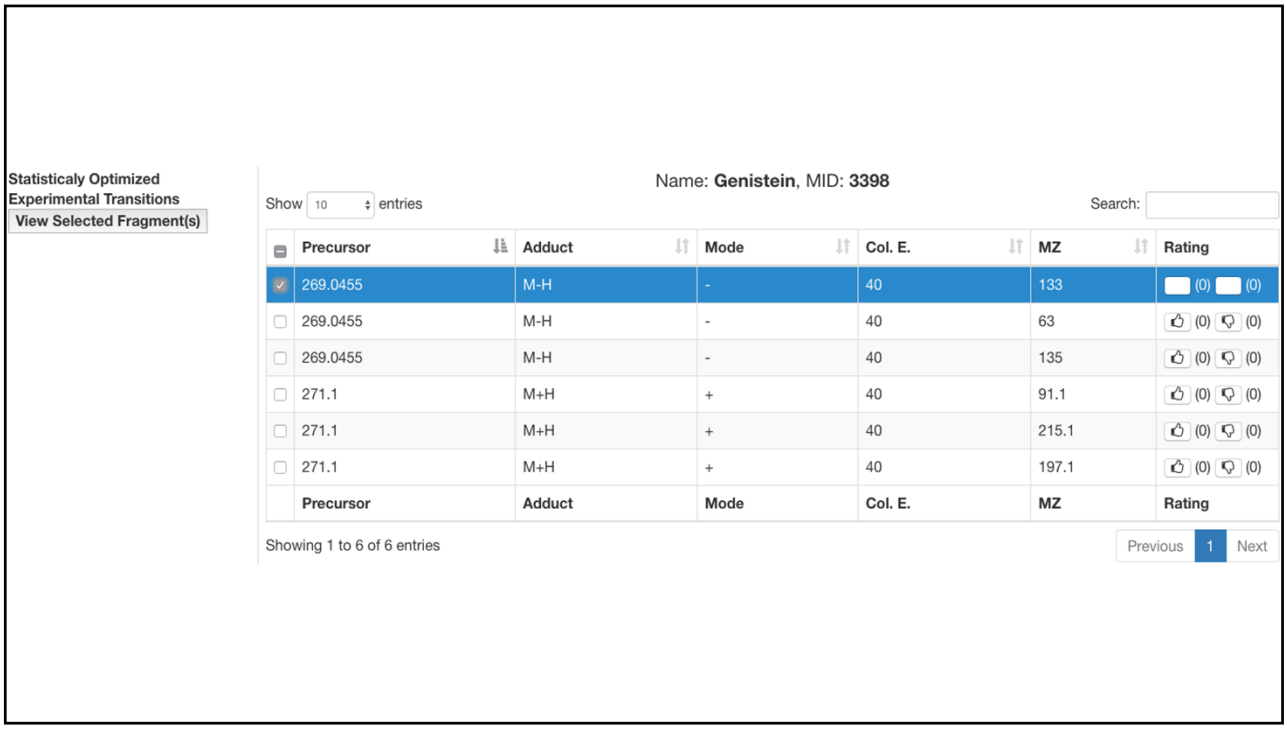
METLIN ID	Mass	Name	Formula	CAS	KEGG	MRM	MS/MS	Structure
3398	270.05282343	Genistein	C15H10O5	446-72-0	C06563	View	View	

METLIN ID	Mass	Name	Formula	CAS	KEGG	MRM	MS/MS	Structure

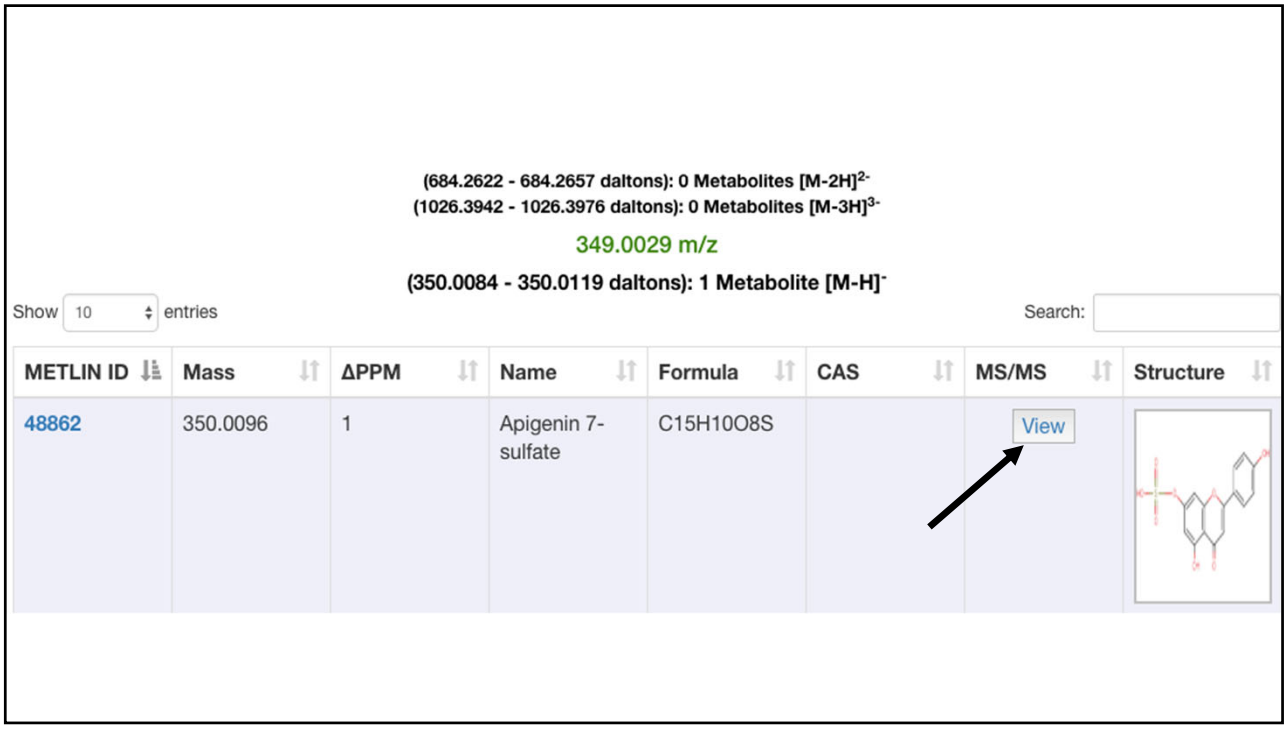
Showing 1 to 1 of 1 entries

Previous 1 Next

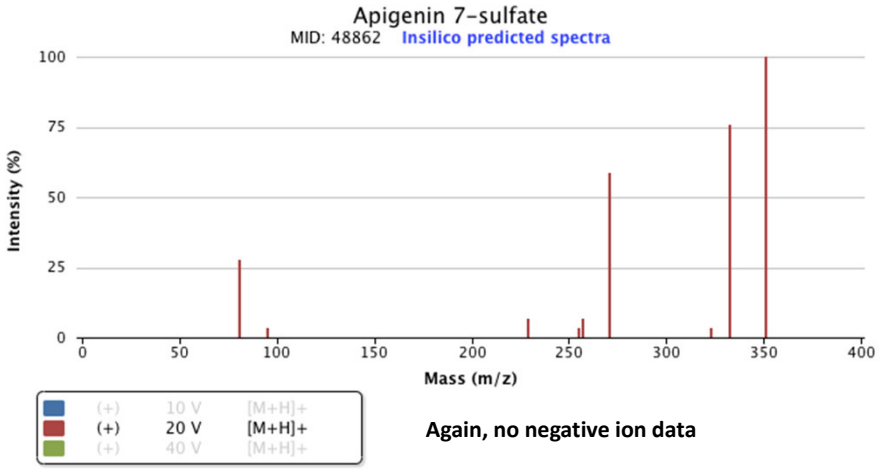
26


Statisticaly Optimized
Experimental Transitions
View Selected Fragment(s)
Name: Genistein, MID: 3398
Show 10 entries
Search:
Precursor
Adduct
Mode
Col. E.
MZ
Rating
269.0455 M-H - 40 133 (0) (0)
269.0455 M-H - 40 63 (0) (0)
269.0455 M-H - 40 135 (0) (0)
271.1 M+H + 40 91.1 (0) (0)
271.1 M+H + 40 215.1 (0) (0)
271.1 M+H + 40 197.1 (0) (0)
Showing 1 to 6 of 6 entries
Previous 1 Next

27


(684.2622 - 684.2657 daltons): 0 Metabolites [M-2H]2-
(1026.3942 - 1026.3976 daltons): 0 Metabolites [M-3H]3-
349.0029 m/z
(350.0084 - 350.0119 daltons): 1 Metabolite [M-H]-
Show 10 entries
Search:
METLIN ID
Mass
ΔPPM
Name
Formula
CAS
MS/MS
Structure
48862
350.0096
1
Apigenin 7-sulfate
C15H10O8S
View

28



Again, no negative ion data

29

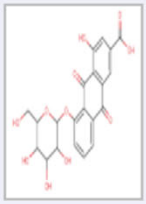
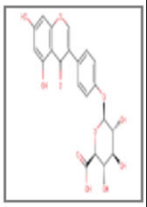
445.0773 m/z

(446.0824 - 446.0868 daltons): 12 Metabolites [M-H]⁻

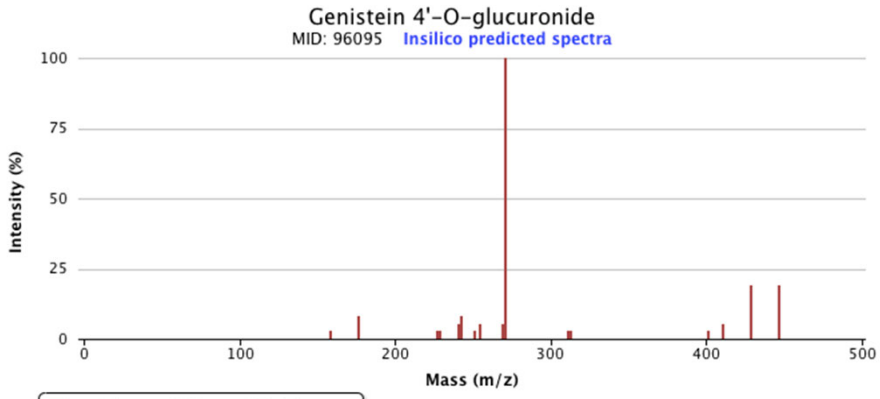
Show 10 entries Search:

METLIN ID	Mass	ΔPPM	Name	Formula	CAS	MS/MS	Structure
48550	446.0849	0	5,7,2'-Trihydroxyflavone 7-glucuronide	C21H18O11		View	
48775	446.0849	0	Apigenin 7-glucuronide	C21H18O11		View	

30

88791	446.0849	0	Glucorhein	C21H18O11	34298-86-7	View	
96095	446.0849	0	Genistein 4'-O-glucuronide	C21H18O11		View	

31



Again, no negative ion data

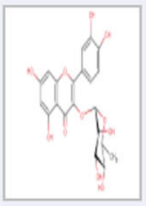
32

This is really the $^{13}C_2$-isotope ion of *m/z* 445.077

447.0932 m/z

(448.0982 - 448.1027 daltons): 74 Metabolites [M-H]⁻

Show 10 entries Search:

METLIN ID	Mass	ΔPPM	Name	Formula	CAS	MS/MS	Structure
43747	448.1006	0	Quercitrin	C21H20O11	522-12-3	View	

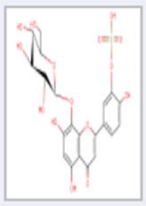
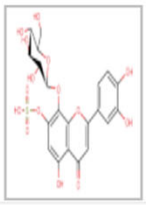
33

525.0333 m/z

(526.038 - 526.0432 daltons): 0 Metabolites [M-H]⁻

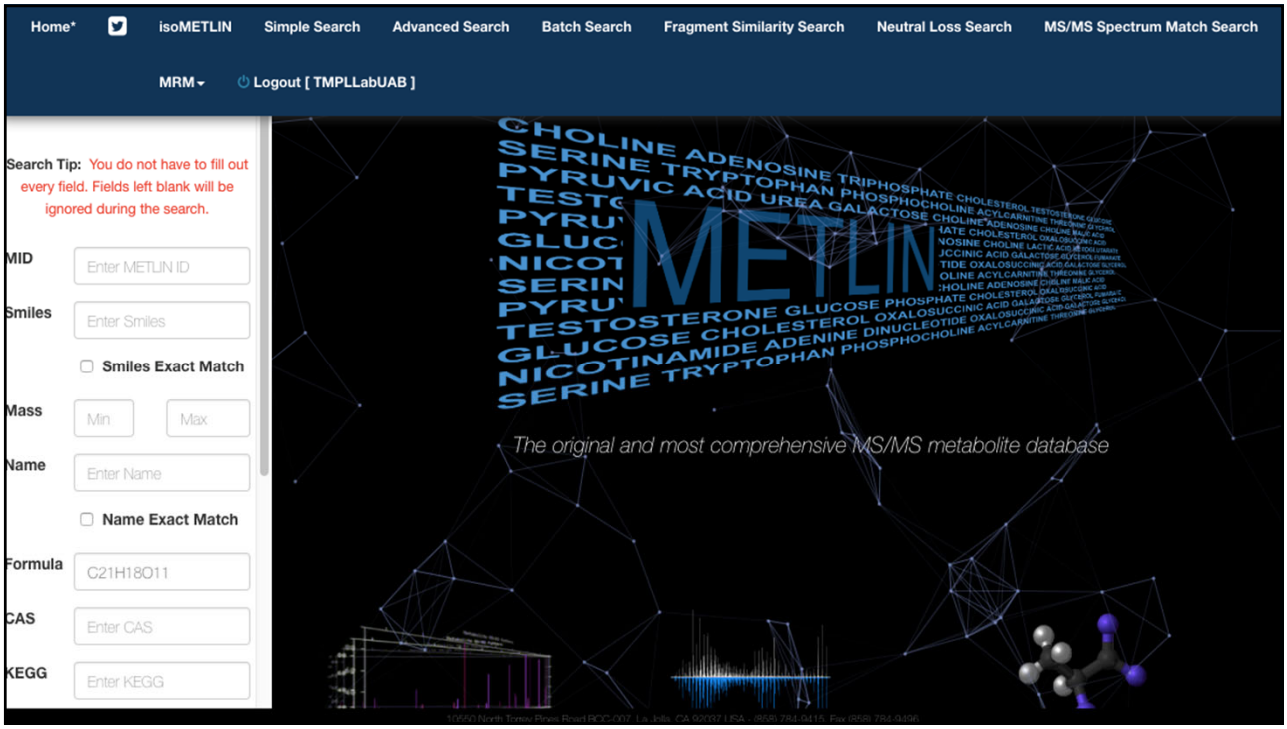
(544.0491 - 544.0543 daltons): 5 Metabolites [M-H2O-H]⁻

Show 10 entries Search:

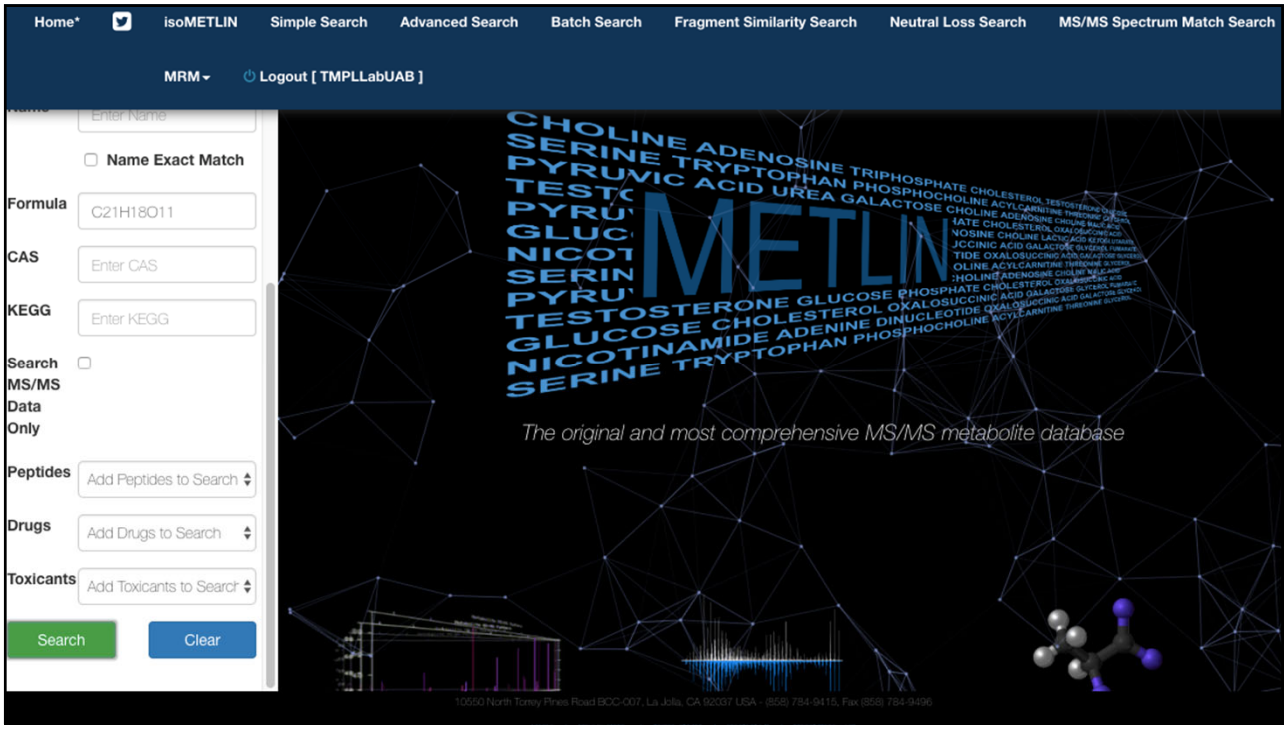
METLIN ID	Mass	ΔPPM	Name	Formula	CAS	MS/MS	Structure
49817	544.0523	1	8-Hydroxyluteolin 8-glucoside-3'-sulfate	C21H20O15S		View	
49819	544.0523	1	Hypolaetin 7-sulfate-8-glucoside	C21H20O15S		View	

None of the five records are of genistein β-glucuronide sulfate

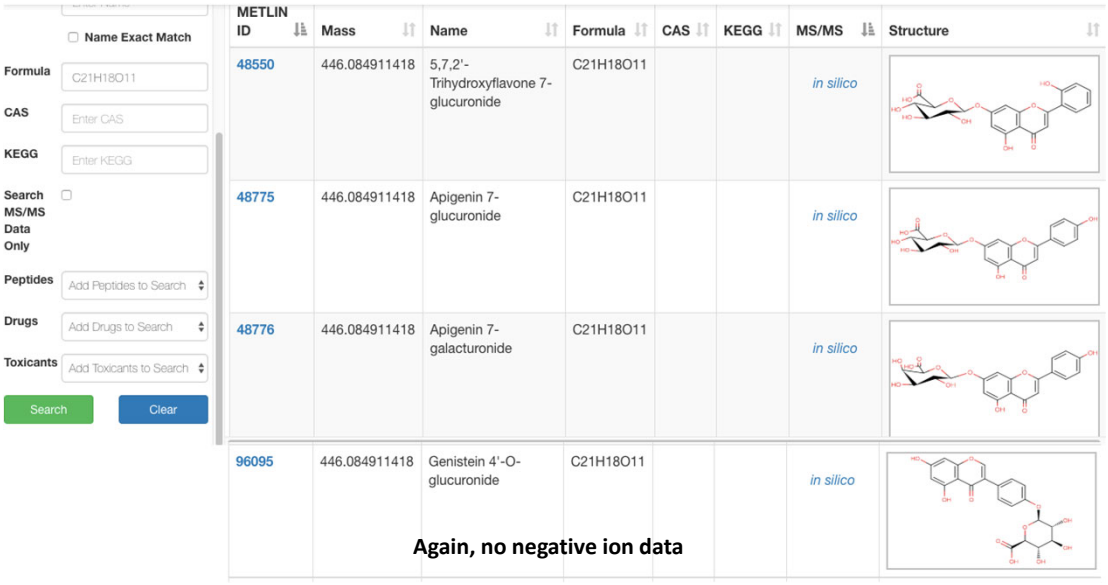
34


Home*
isoMETLIN
Simple Search
Advanced Search
Batch Search
Fragment Similarity Search
Neutral Loss Search
MS/MS Spectrum Match Search
MRM
Logout [TMPLLabUAB]
Search Tip: You do not have to fill out every field. Fields left blank will be ignored during the search.
MID
Enter METLIN ID
Smiles
Enter Smiles
Smiles Exact Match
Mass
Min
Max
Name
Enter Name
Name Exact Match
Formula
C21H18O11
CAS
Enter CAS
KEGG
Enter KEGG
METLIN
The original and most comprehensive MS/MS metabolite database

35


Home*
isoMETLIN
Simple Search
Advanced Search
Batch Search
Fragment Similarity Search
Neutral Loss Search
MS/MS Spectrum Match Search
MRM
Logout [TMPLLabUAB]
Name Exact Match
Formula
C21H18O11
CAS
Enter CAS
KEGG
Enter KEGG
Search MS/MS Data Only
Peptides
Add Peptides to Search
Drugs
Add Drugs to Search
Toxicants
Add Toxicants to Search
Search
Clear
METLIN
The original and most comprehensive MS/MS metabolite database

36



37

Class exercise due Friday, Feb 28

- **The ordering of the features was based on the control group**
- **Go back to the 5<RT<25 min spreadsheet**
  - Copy it to a new spreadsheet and order it according the areas in the genistein group (largest to smallest)
  - Divide the data into two equal parts and transfer them to two separate sheets – the first half of most intense ions (GenCon_1_1) and the second half the remaining less intense ions (GenCon_1_2)
  - Make .csv files for each sample for each half and place them in the folders
  - Make .zip files of the two folders and rename them GenCon_1.zip and GenCon_2.zip
- **Perform stats analysis with Metaboanalyst**
  - Make the figures at the beginning of this slide set (volcano plot, PCA, PLS-DA, VIP and Heatmaps).
  - Download the data and generate a report.
  - Download the download.zip file and unzip it

38

More on the class exercise

- **Locate the plsda_vip file and open it**
- **Parse the m_z/RT column A (insert a new column B first)**
  - Highlight column A and go to the Sort tab – scroll down and select the text to column option
  - Select Other and add a "/" in the box
  - You should see that the concatenated info separates into two columns
  - Once finished, highlight all the data and sort on Comp 1 (largest to smallest).
  - Scroll down until the last row is 3 or greater.
  - Insert an empty row
  - Highlight all the rows above the empty row and sort according to column A (*m/z,* smallest to largest)
  - Identify the all ^{12}C-ions and their $^{13}C_n$-isotope pairs
- **Copy the *m/z* values of ions with comp 1 >3 into METLIN**
  - Use negative ions, [M-H], [M-H_2O-H], [2M-H], and [3M-H] and 5 ppm to do a batch search
  - Identify as many ions as you can

39