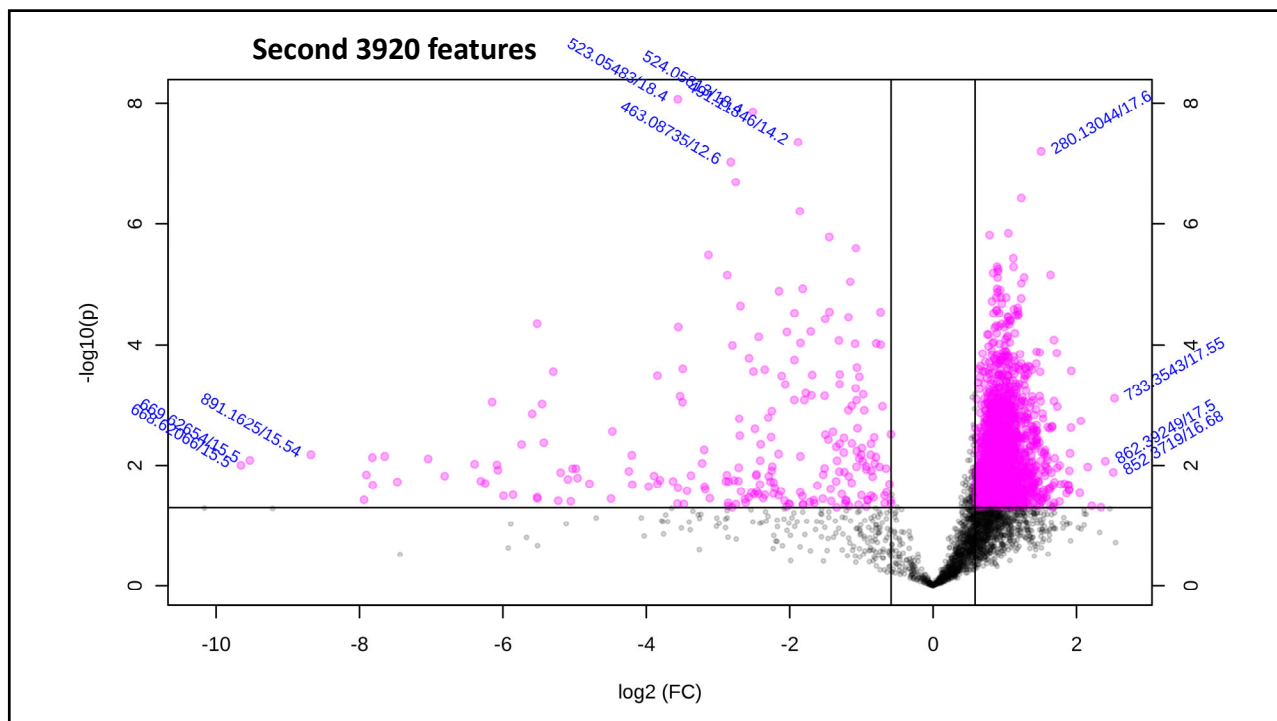
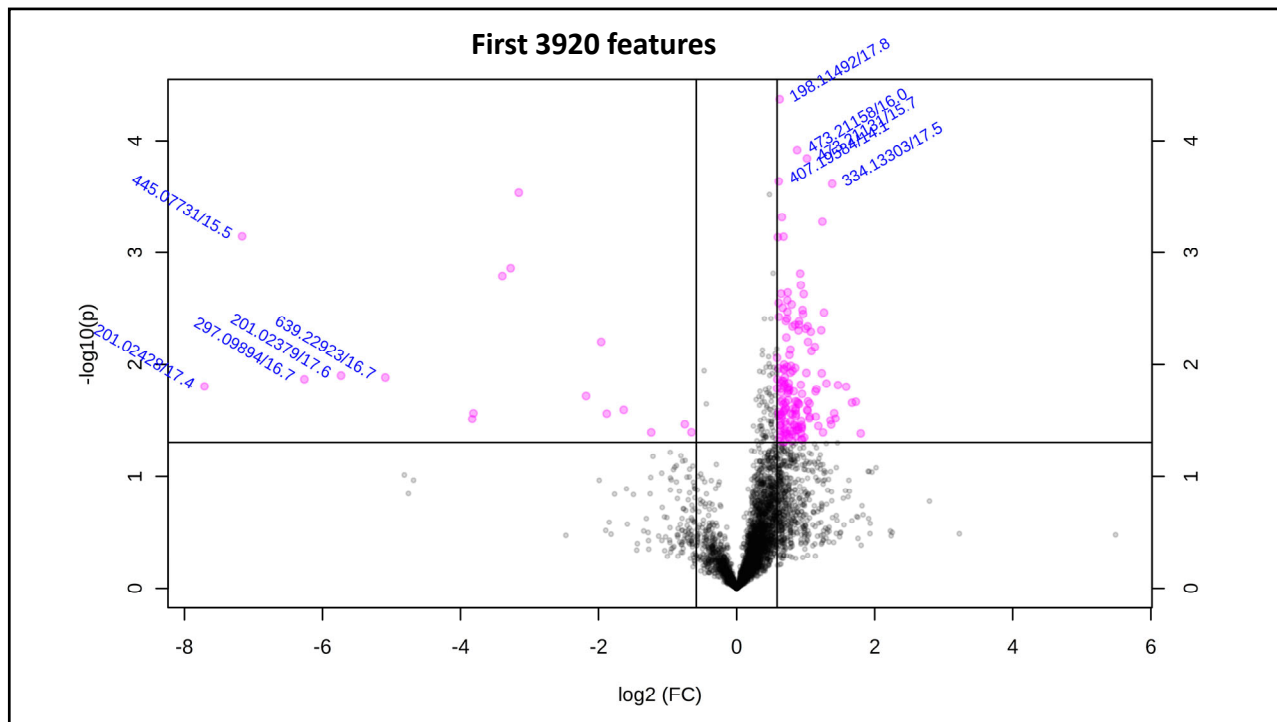


# Dividing up the data

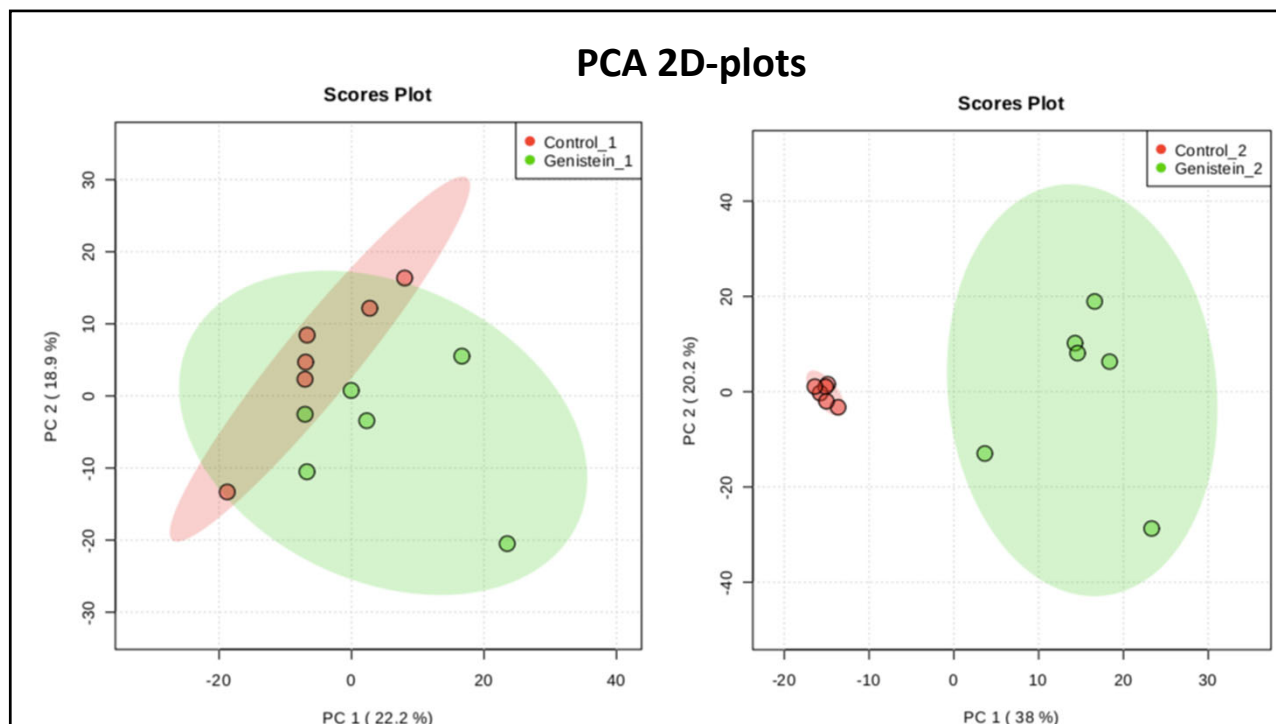
Stephen Barnes

## Analysis of the control-genistein data

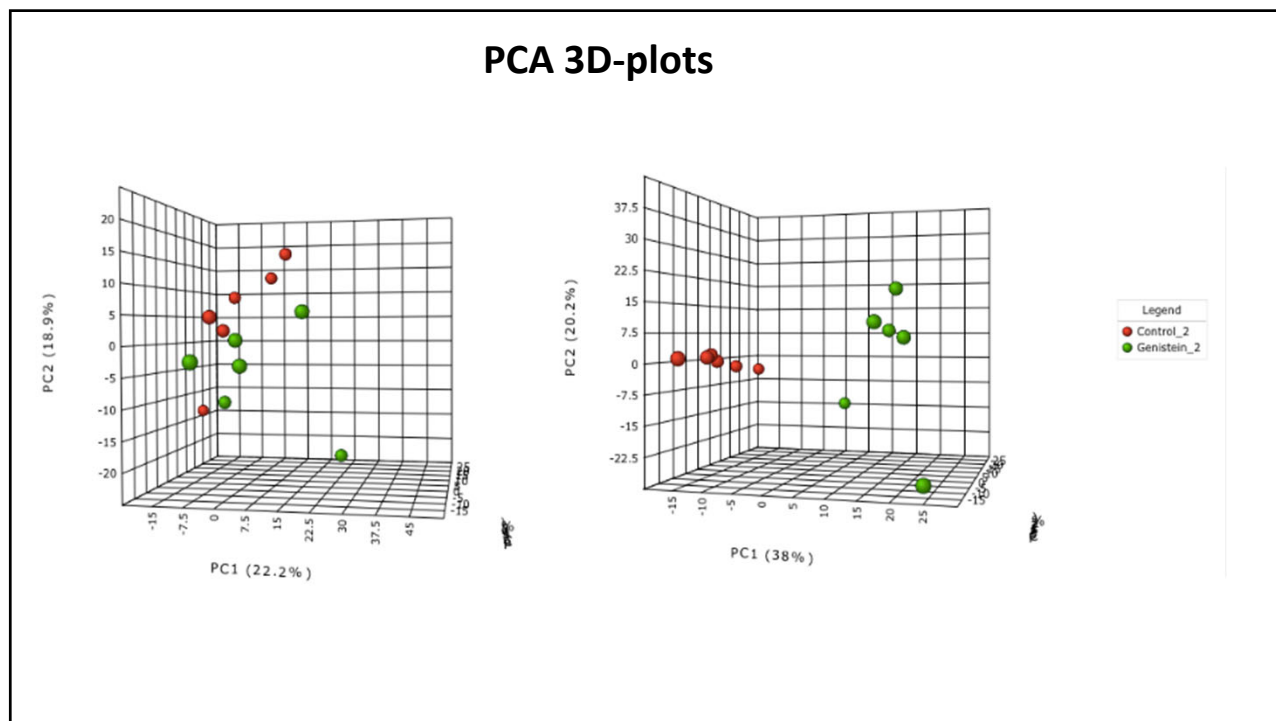
- **The Excel download from the XCMS analysis contained 7840 ion features**
  - Too many to load them onto MetaboAnalyst (restricted to 5000 features)
- **The data on the 5<RT<25 min sheet was ordered (highest to lowest) according to the values in the control mean column**
  - These ordered data were 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

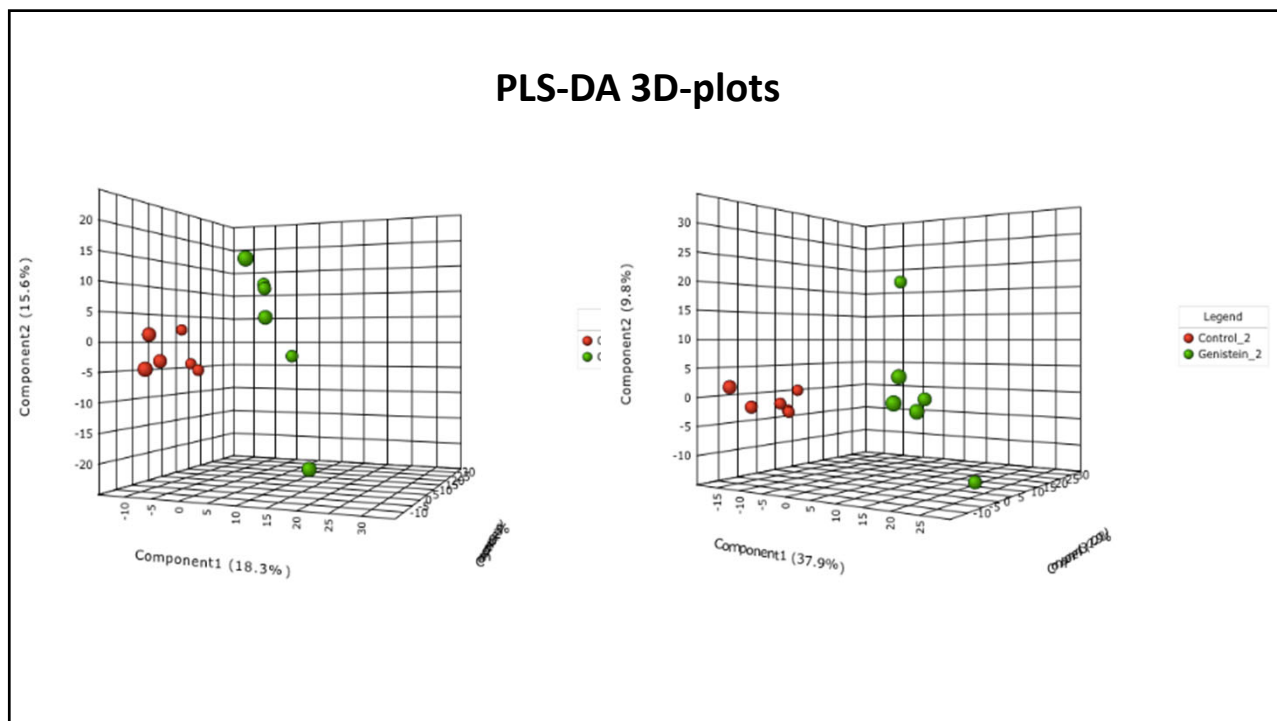
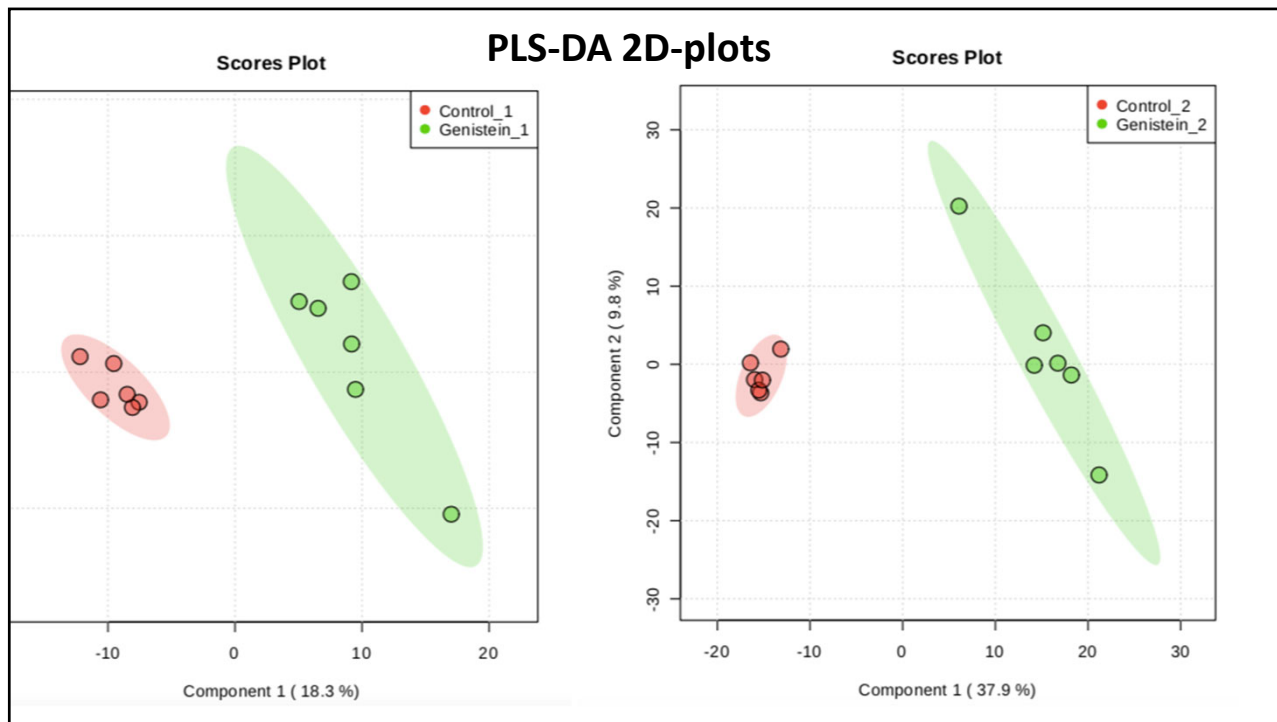


## PCA 2D-plots



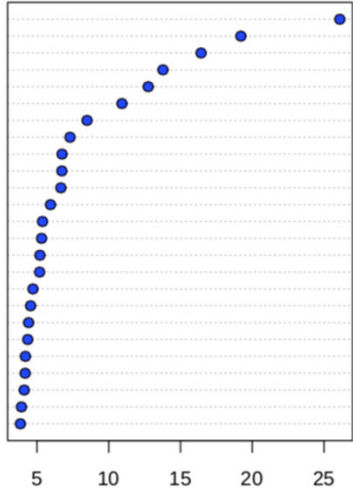
## PCA 3D-plots



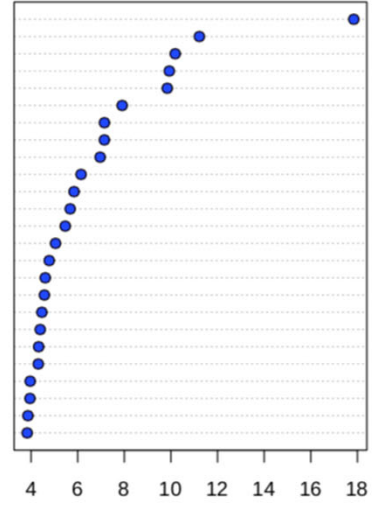


### VIP plots

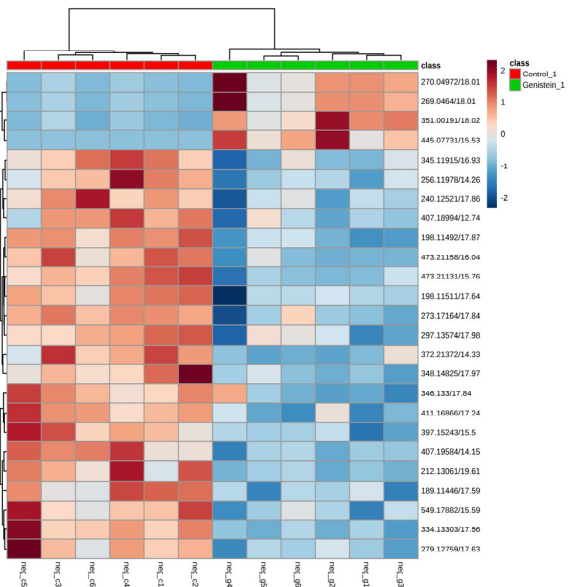
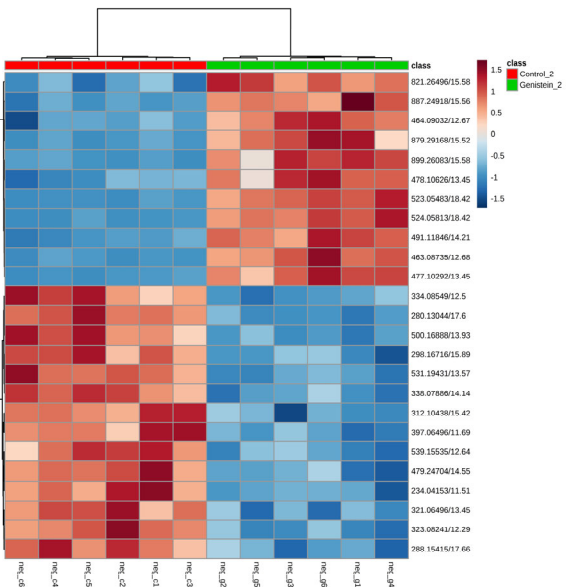
349.00285/18.0  
 201.02428/17.4  
 445.07731/15.5  
 297.09894/16.7  
 201.02379/17.6  
 350.00603/18  
 269.0464/18.01  
 639.22923/16.7  
 283.11787/14.8  
 341.12466/16.7  
 351.00191/18.0  
 525.03334/13.5  
 427.1796/18.82  
 525.03352/14.1  
 407.10162/12.6  
 319.1225/18.39  
 325.18459/21.7  
 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



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



### Heat Maps



m/z	RT		m/z	RT	
112.9885	26.23		202.0273	17.45	
201.0238	17.64		203.0209	17.39	
201.0243	17.40		262.0139	13.51	
269.0136	14.06		269.0462	15.53	
269.0464	18.01		269.0465	20.50	
270.0497	18.01		298.1022	16.79	
283.0832	14.73		351.0178	17.70	
283.1179	14.87		352.0046	17.99	
291.0915	17.87		373.1142	12.75	
291.0916	17.06		425.0347	17.43	
297.0989	16.76		433.0774	10.67	
319.1225	18.39		445.0770	16.31	
325.1845	20.70		446.0814	15.53	
325.1846	21.71		447.0915	15.52	
341.1247	16.74		448.0947	15.53	
349.0029	18.02		449.1089	15.78	
350.0060	18.00		461.1087	19.25	
351.0019	18.02		463.0874	12.68	
380.1745	16.77		477.1029	13.45	
404.1920	13.32		511.0543	14.22	
407.1016	12.69		523.0548	18.42	
411.1269	11.12		526.0369	14.14	
415.1970	19.07		526.0370	13.51	
425.1812	16.68		595.2033	16.78	
427.1796	18.82		596.2067	16.78	
429.0494	11.88		597.2168	16.79	
441.1585	15.09		621.1097	12.57	
443.1743	13.65		640.2331	16.77	
445.0773	15.53		668.1188	15.53	
447.0932	16.18		668.6207	15.53	
525.0333	13.51		669.1240	15.53	
525.0335	14.14		675.3013	16.78	
639.2292	16.76		891.1625	15.54	
683.2553	16.74		892.1656	15.54	
823.2608	11.12		893.1741	15.54	
			894.1785	15.54	

**Significant ions from VIP sets  
Comp 1 >3, ordered by  $m/z$**

## 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  $2 \times 668.119 + 2 \times 1.0072 = 1336.238 + 2.014 = 1338.252$
  - Divide  $1338.252$  by  $446.086 = 3$
- $m/z$  668.119 is therefore  $[3M-2H]^{2-}$
- All these ions are from genistein glucuronide

**Now let's go to METLIN**  
 Put METLIN in your browser  
 use your account information

Home\* Metlin IsoMETLIN Sign Up Forgot Password?

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**

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**

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



Home\* isoMETLIN Simple Search Advanced Search Batch Search Fragment Similarity Search Neutral Loss Search MS/MS Spectrum Match Search

MRM Logout [ TEMPLabUAB ]

M+2H  
M-3H  
M+CH<sub>3</sub>COO  
M+F

Accuracy (PPM)  Select ppm = 5

Display  Structure

Peptides

Drugs

Toxicants

Go search

**METLIN**

CHOLINE ADENOSINE TRIPHOSPHATE CHOLESTEROL TESTOSTERONE SERINE TRYPTOPHAN PHOSPHOCHOLINE ACYL-CARNITINE THREONINE GLUCOSE PYRUVIC ACID UREA GALACTOSE CHOLINE ADENOSINE CHOLINE MULTY... TESTOSTERONE GLUCOSE PHOSPHATE CHOLESTEROL OXALOSUC... NICOTINAMIDE ADENINE DINUCLEOTIDE OXALOSUC... SERINE TRYPTOPHAN PHOSPHOCHOLINE ACYL-CARNITINE THREONINE...

The original and most comprehensive MS/MS metabolite database

112.9885 m/z  
(113.9952 - 113.9963 daltons): 1 Metabolite [M-H]<sup>-</sup>

Show  entries Search:

METLIN ID	Mass	ΔPPM	Name	Formula	CAS	MS/MS	Structure
63238	113.9953	4	Acetylenedicarboxylate	C <sub>4</sub> H <sub>2</sub> O <sub>4</sub>	142-45-0	<a href="#">View</a>	

Acetylenedicarboxylate  
MID: 63238 [insilico predicted spectra](#)

Intensity (%)

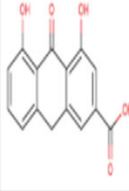
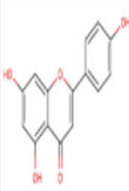
Mass (m/z)

(+) 10 V [M+H]<sup>+</sup>  
 (+) 20 V [M+H]<sup>+</sup>  
 (+) 40 V [M+H]<sup>+</sup>

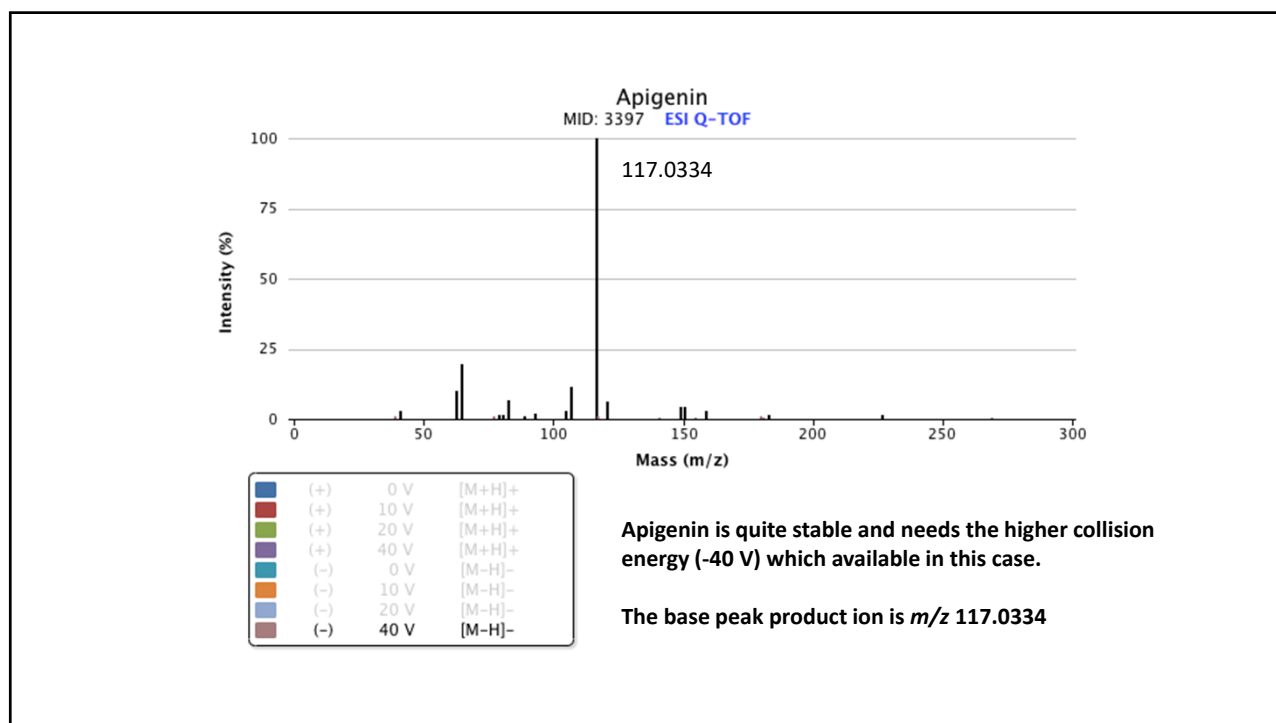
**MSMS spectrum not useful. Would be better if METLIN didn't link negative ions to positive MSMS spectra**

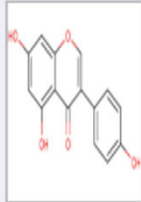
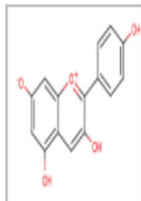
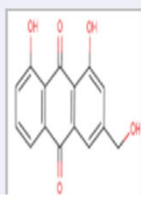
(270.0523 - 270.055 daltons): 29 Metabolites [M-H]<sup>-</sup>

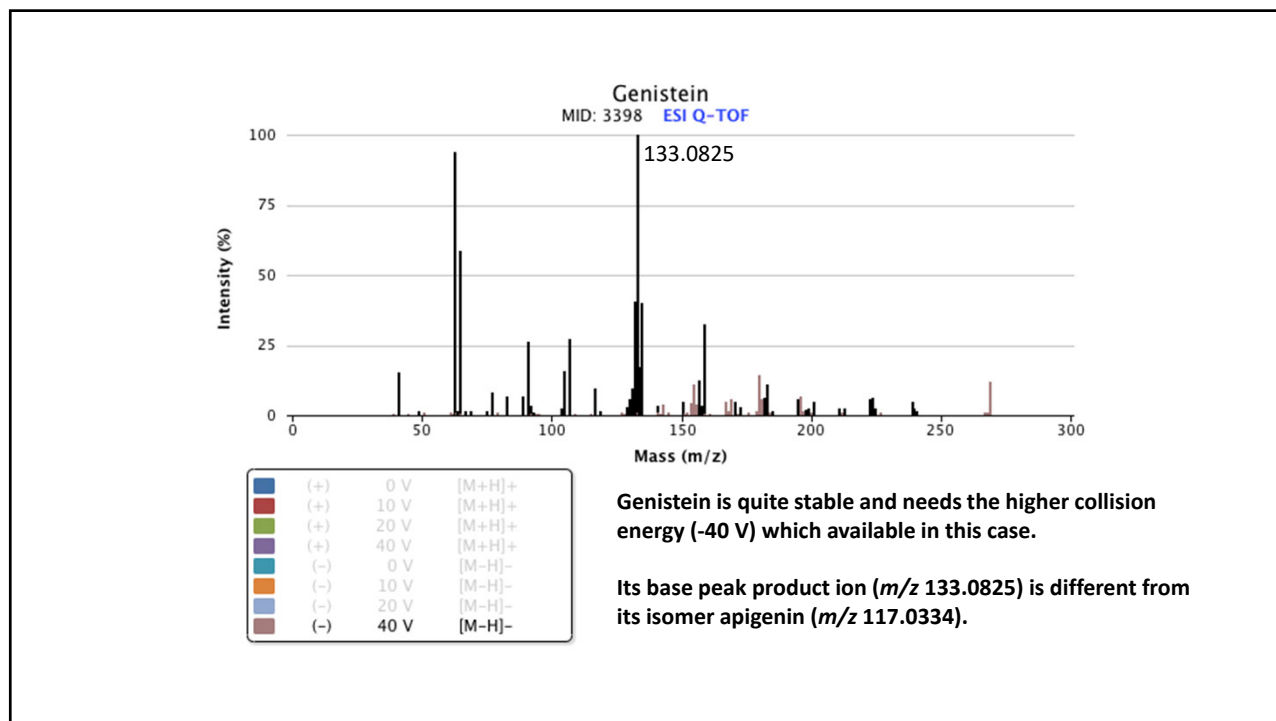
Show 10 entries Search:

METLIN ID	Mass	$\Delta$ PPM	Name	Formula	CAS	MS/MS	Structure
<a href="#">2412</a>	270.0528	3	Rhein-9-anthrone	C <sub>15</sub> H <sub>10</sub> O <sub>5</sub>	480-09-1	<a href="#">View</a>	
<a href="#">3397</a>	270.0528	3	Apigenin	C <sub>15</sub> H <sub>10</sub> O <sub>5</sub>	520-36-5	<a href="#">View</a>	

↗



3398	270.0528	3	Genistein	C <sub>15</sub> H <sub>10</sub> O <sub>5</sub>	446-72-0	<a href="#">View</a>	
3399	270.0528	3	Pelargonidin	C <sub>15</sub> H <sub>10</sub> O <sub>5</sub>	7690-51-9	<a href="#">View</a>	
41039	270.0528	3	Aloe-emodin	C <sub>15</sub> H <sub>10</sub> O <sub>5</sub>	NA	<a href="#">View</a>	



Home\* isoMETLIN Simple Search Advanced Search Batch Search Fragment Similarity Search Neutral Loss Search MS/MS Spectrum Match Search

MRM Logout [ TMPLLabUAB ]

M-2F  
M-3H  
M+CH3  
M+F

METLIN MRM  
MRM Upload  
MRM Download

		270.0528	3	Genistein	C15H10O5	446-72-0	<a href="#">View</a>		
Accuracy (PPM)	5	3399	270.0528	3	Pelargonidin	C15H10O5	7690-51-9	<a href="#">View</a>	
Display Structure	<input checked="" type="checkbox"/>								
Peptides	Remove Peptides from								
Drugs	Remove Drugs from St	41039	270.0528	3	Aloe-emodin	C15H10O5	NA	<a href="#">View</a>	
Toxicants	Remove Toxicants from								

Search Clear

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

Smiles

Smiles Exact Match

Mass

Name

Name Exact Match

Formula

CAS

METLIN

The original and most comprehensive MS/MS metabolite database

Home\* isoMETLIN Simple Search Advanced Search Batch Search Fragment Similarity Search Neutral Loss Search MS/MS Spectrum Match Search

MRM Logout [ TMPLLabUAB ]

Enter Name

Name Exact Match

Formula

CAS

KEGG

Search

MS/MS Data Only

Peptides

Drugs

Toxicants

**METLIN**

CHOLINE ADENOSINE TRIPHOSPHATE CHOLESTEROL TESTOSTERONE GLUCOSE PHOSPHATE CHOLESTEROL  
 SERINE ADENOSINE TRIPHOSPHATE CHOLESTEROL TESTOSTERONE GLUCOSE PHOSPHATE CHOLESTEROL  
 PYRUVIC ACID UREA GALACTOSE CHOLINE ADENOSINE TRIPHOSPHATE CHOLESTEROL  
 TESTOSTERONE GLUCOSE PHOSPHATE CHOLESTEROL  
 PYRUVIC ACID UREA GALACTOSE CHOLINE ADENOSINE TRIPHOSPHATE CHOLESTEROL  
 GLUCOSE PHOSPHATE CHOLESTEROL  
 NICOTINAMIDE ADENINE DINUCLEOTIDE  
 SERINE TRYPTOPHAN PHOSPHOCHOLINE ACYLGLUTAMINE THREONINE SUCROSE

The original and most comprehensive MS/MS metabolite database

METLIN ID	Mass	Name	Formula	CAS	KEGG	MRM	MS/MS	Structure
3398	270.05282343	Genistein	C15H10O5	446-72-0	C06563	<a href="#">View</a>	<a href="#">View</a>	

Showing 1 to 1 of 1 entries

Previous [1](#) Next

Statistically Optimized  
Experimental Transitions

[View Selected Fragment\(s\)](#)

Name: **Genistein**, MID: 3398

Show 10 entries

Search:

<input type="checkbox"/>	Precursor	Adduct	Mode	Col. E.	MZ	Rating
<input checked="" type="checkbox"/>	269.0455	M-H	-	40	133	<input type="checkbox"/> (0) <input type="checkbox"/> (0)
<input type="checkbox"/>	269.0455	M-H	-	40	63	<input type="checkbox"/> (0) <input type="checkbox"/> (0)
<input type="checkbox"/>	269.0455	M-H	-	40	135	<input type="checkbox"/> (0) <input type="checkbox"/> (0)
<input type="checkbox"/>	271.1	M+H	+	40	91.1	<input type="checkbox"/> (0) <input type="checkbox"/> (0)
<input type="checkbox"/>	271.1	M+H	+	40	215.1	<input type="checkbox"/> (0) <input type="checkbox"/> (0)
<input type="checkbox"/>	271.1	M+H	+	40	197.1	<input type="checkbox"/> (0) <input type="checkbox"/> (0)
<input type="checkbox"/>	Precursor	Adduct	Mode	Col. E.	MZ	Rating

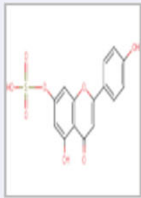
Showing 1 to 6 of 6 entries

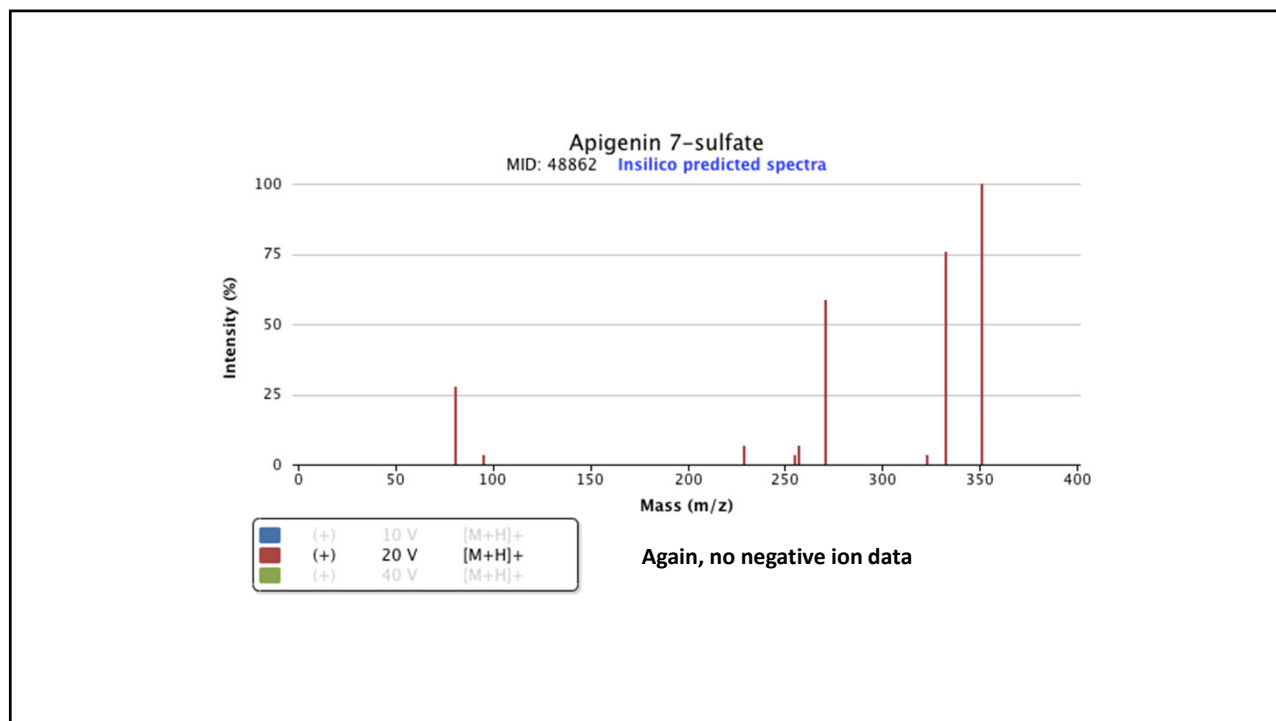
Previous **1** Next

(684.2622 - 684.2657 daltons): 0 Metabolites [M-2H]<sup>2-</sup>  
 (1026.3942 - 1026.3976 daltons): 0 Metabolites [M-3H]<sup>3-</sup>  
**349.0029 m/z**  
 (350.0084 - 350.0119 daltons): 1 Metabolite [M-H]<sup>-</sup>

Show 10 entries

Search:

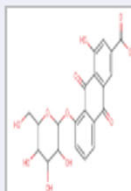
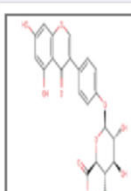
METLIN ID	Mass	$\Delta$ PPM	Name	Formula	CAS	MS/MS	Structure
48862	350.0096	1	Apigenin 7-sulfate	C <sub>15</sub> H <sub>10</sub> O <sub>8</sub> S		<a href="#">View</a>	

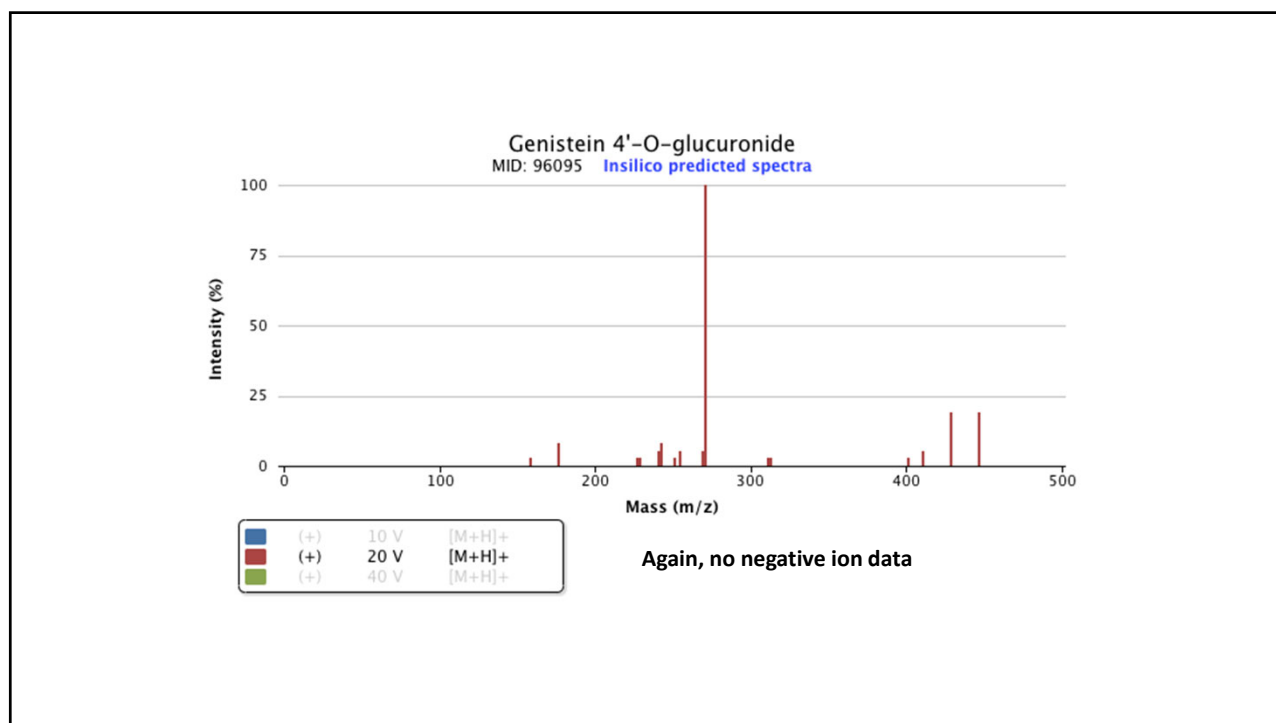


**445.0773 m/z**  
(446.0824 - 446.0868 daltons): 12 Metabolites [M-H]<sup>-</sup>

Show  entries Search:

METLIN ID	Mass	ΔPPM	Name	Formula	CAS	MS/MS	Structure
<a href="#">48550</a>	446.0849	0	5,7,2'-Trihydroxyflavone 7-glucuronide	C <sub>21</sub> H <sub>18</sub> O <sub>11</sub>		<a href="#">View</a>	
<a href="#">48775</a>	446.0849	0	Apigenin 7-glucuronide	C <sub>21</sub> H <sub>18</sub> O <sub>11</sub>		<a href="#">View</a>	

88791	446.0849	0	Glucorhein	C <sub>21</sub> H <sub>18</sub> O <sub>11</sub>	34298-86-7	<a href="#">View</a>	
96095	446.0849	0	Genistein 4'-O-glucuronide	C <sub>21</sub> H <sub>18</sub> O <sub>11</sub>		<a href="#">View</a>	





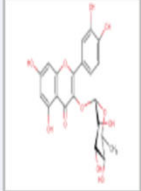
This is really the  $^{13}\text{C}_2$ -isotope ion of  $m/z$  445.077

↓

447.0932  $m/z$

(448.0982 - 448.1027 daltons): 74 Metabolites  $[\text{M-H}]^-$

Show  entries Search:

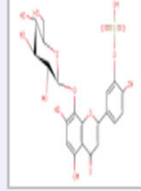
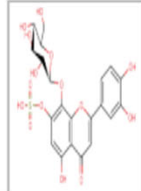
METLIN ID ↓	Mass ↓	$\Delta\text{PPM}$ ↓	Name ↓	Formula ↓	CAS ↓	MS/MS ↓	Structure ↓
43747	448.1006	0	Quercitrin	C <sub>21</sub> H <sub>20</sub> O <sub>11</sub>	522-12-3	<a href="#">View</a>	

525.0333  $m/z$

(526.038 - 526.0432 daltons): 0 Metabolites  $[\text{M-H}]^-$

(544.0491 - 544.0543 daltons): 5 Metabolites  $[\text{M-H}_2\text{O-H}]^-$

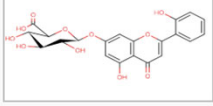
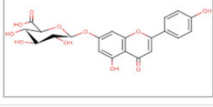
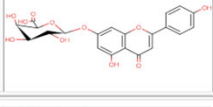
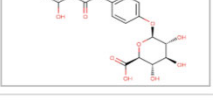
Show  entries Search:

METLIN ID ↓	Mass ↓	$\Delta\text{PPM}$ ↓	Name ↓	Formula ↓	CAS ↓	MS/MS ↓	Structure ↓
49817	544.0523	1	8-Hydroxyluteolin 8-glucoside-3'-sulfate	C <sub>21</sub> H <sub>20</sub> O <sub>15</sub> S		<a href="#">View</a>	
49819	544.0523	1	Hypolaetin 7-sulfate-8-glucoside	C <sub>21</sub> H <sub>20</sub> O <sub>15</sub> S		<a href="#">View</a>	

**None of the five records are of genistein  $\beta$ -glucuronide sulfate**



Name Exact Match  
 Formula   
 CAS   
 KEGG   
 Search MS/MS Data Only  
 Peptides   
 Drugs   
 Toxicants

METLIN ID	Mass	Name	Formula	CAS	KEGG	MS/MS	Structure
48550	446.084911418	5,7,2'-Trihydroxyflavone 7-glucuronide	C21H18O11			<i>in silico</i>	
48775	446.084911418	Apigenin 7-glucuronide	C21H18O11			<i>in silico</i>	
48776	446.084911418	Apigenin 7-galacturonide	C21H18O11			<i>in silico</i>	
96095	446.084911418	Genistein 4'-O-glucuronide	C21H18O11			<i>in silico</i>	

**Again, no negative ion data**

## Class exercise due Friday, Feb 22

- **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

## More on the class exercise

- **Locate the plstda\_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  $^{12}\text{C}$ -ions and their  $^{13}\text{C}_n$ -isotope pairs
- **Copy the  $m/z$  values of ions with comp 1 >3 into METLIN**
  - Use negative ions, [M-H], [M-H<sub>2</sub>O-H], [2M-H], and [3M-H] and 5 ppm to do a batch search
  - Identify as many ions as you can