evropský sociální fond v ČR
EVROPSKÁ UNIE
MINISTERSTVO ŠKOLSTVÍ, MLÁDEŽE A TĚLOVÝCHOVY
OP Vzdělávání pro konkurenceschopnost
INVESTICE DO ROZVOJE VZDĚLÁVÁNÍ

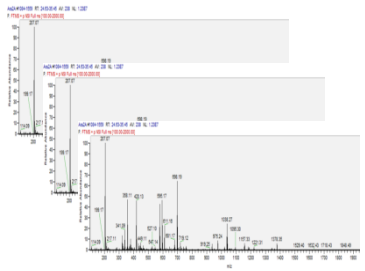
Principles and Practices on analyzing metabolomics data

Dr. Xiaoliang Sun

Laboratory of Ecological Plant Physiology

Tato akce se koná v rámci projektu:
Vybudování vědeckého týmu environmentální metabolomiky a ekofyziologie a jeho zapojení do mezinárodních sítí (ENVIMET; r.č. **CZ.1.07/2.3.00/20.0246**) realizovaného v rámci Operačního programu Vzdělávání pro konkurenceschopnost.

Outline

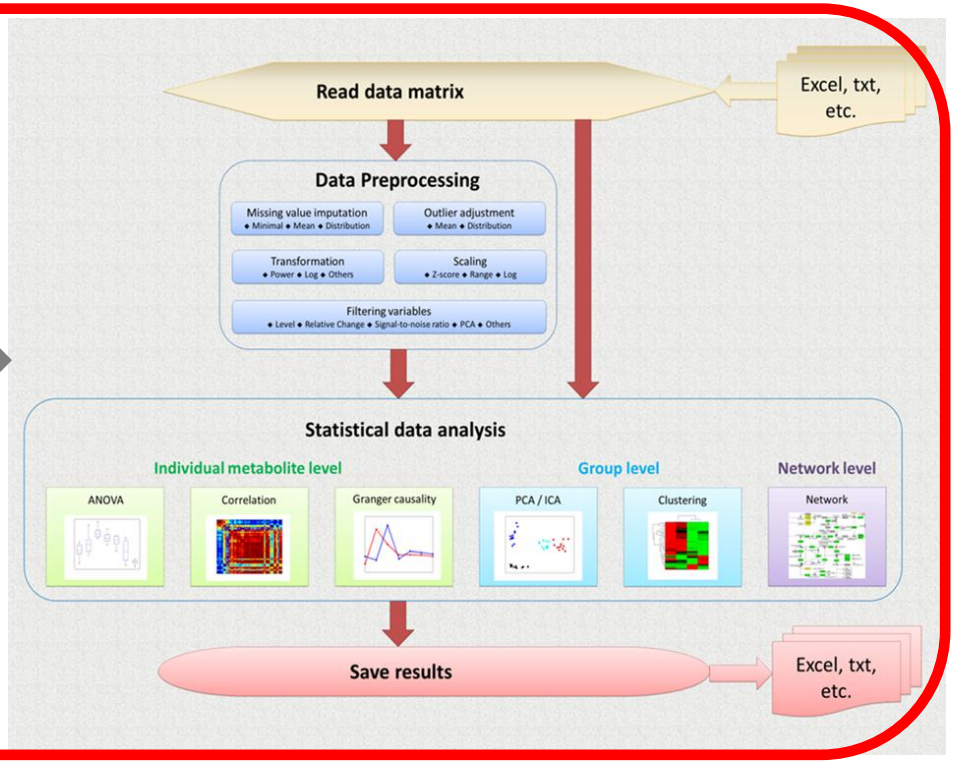


A	B	C	D	E	F
m/z	Intensity	Relative	Theo. Mass	Delta (ppm)	Composition
448.11559	25978.6	0.86	448.11578	-0.41	C11 H22 O14 N5
			448.11527	0.72	C25 H20 O8
			448.11526	0.73	C24 H14 O3 N7
449.10783	188211.2	6.2	449.10783	-0.01	C20 H15 O6 N7
			449.10784	-0.02	C21 H21 O11
594.18977	9240.6	0.3	594.18977	-0.01	C36 H26 O5 N4
			594.19028	-0.86	C22 H28 O11 N9
			594.19028	-0.87	C23 H34 O16 N2
594.35642	8594.1	0.28	594.35644	-0.04	C37 H46 O3 N4
			594.35695	-0.89	C23 H48 O9 N9
			594.35696	-0.9	C24 H54 O14 N2
595.14474	71219.1	2.35	595.14462	0.21	C30 H27 O13
			595.14461	0.22	C29 H21 O8 N7
			595.14512	-0.64	C16 H29 O19 N5

Name
Aconitic acid, cis-
Alanine
Alanine-, beta
Ascorbic acid
Aspartic acid
Citric acid
Dehydroascorbic acid
Disaccharid
Ethanolamine
Fructose

Today's topic:

Name	Control	3D	5D	3G	5G
Aconitic acid, cis-	5.327144535	4.77204931	4.75735824	5.41641324	5.03003
Alanine	5.768923669	5.31855579	5.42984623	6.0864222	6.705485
Alanine-, beta	5.18895195	4.8628535	4.94017365	5.35239817	5.535512
Ascorbic acid	5.355646654	5.33068737	5.54944019	5.45579956	5.716773
Aspartic acid	5.835499666	6.07170371	6.73390743	6.14484235	5.97393
Citric acid	7.470201132	7.29674425	7.25328882	7.5181206	7.303258
Dehydroascorbic acid	6.450444957	6.42382375	6.8047005	6.58694452	6.465055
Disaccharid	4.423030873	4.05857819	4.32859405	4.4992042	4.078352
Ethanolamine	7.53389473	7.46042205	7.51635012	7.57852577	7.445617
Fructose	6.568186542	6.41723942	6.97633475	6.54992489	6.457776



Features of metabolomics data

- High dimensionality, i.e., many compounds, treatments
- Data are usually not perfect, e.g., missing values, outliers, ranging in many magnitudes
- Interact in pathways and network
- Limited identification for secondary metabolites

	Condition 1				Condition 2				Condition 3				…	Condition n			
	R 1	R 2	…	R p	R 1	R 2	…	R p	R 1	R 2	…	R p		R 1	R 2	…	R p
Metabolite 1	x	x	x	x	x	x	x	x	x	x	x	x	…	x	x	x	x
Metabolite 2	x	x	x	x	x	x	x	x	x	x	x	x	…	x	x	x	x
Metabolite 3	x	x	x	x	x	x	x	x	x	x	x	x	…	x	x	x	x
…	…	…	…	…	…	…	…	…	…	…	…	…	…	…	…	…	…
Metabolite m	x	x	x	x	x	x	x	x	x	x	x	x	…	x	x	x	x

From this page on, I use “variables” to denote metabolites, and “conditions” to denote treatments (e.g., control, cold stress, etc.)

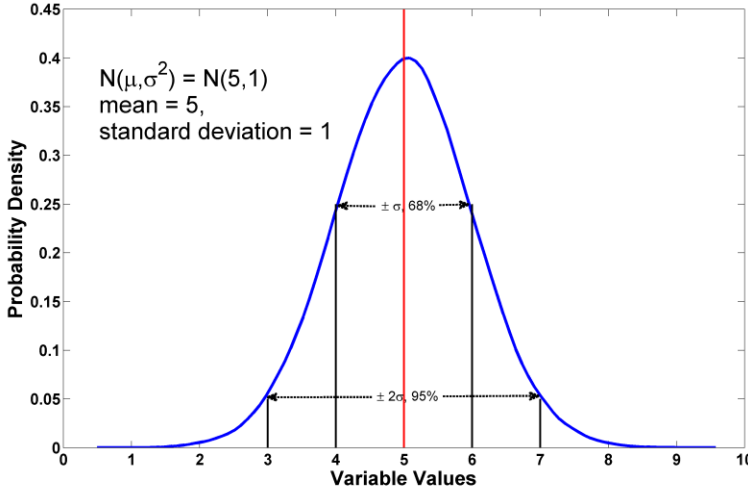
Principles and methods to analyze metabolomics data

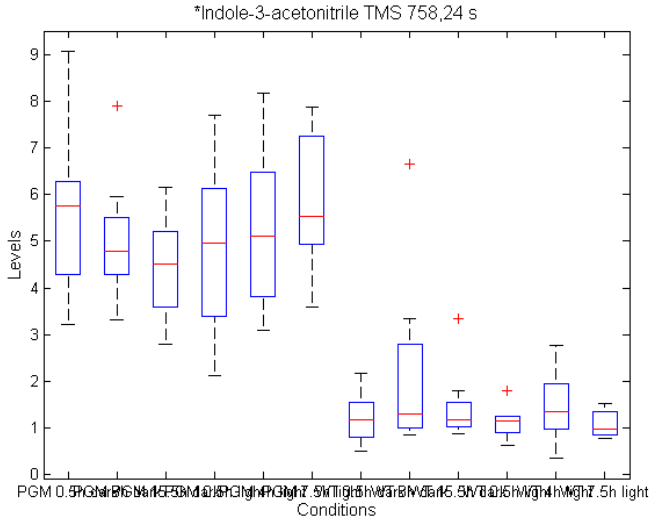
- Individual metabolite level (Univariate analysis)
  - To detect changes under different conditions
  - To study pairwise relationships between metabolites
- Metabolite group level (Multivariate analysis)
  - To study metabolite group differences relating to conditions
  - To study patterns of metabolites profiles across all conditions
  - To find out which factors cause differences
  - To find out which metabolites are important, such as biomarkers
- Network level (Graph analysis)
  - Mapping identified metabolites to pathway databases
  - Inferring metabolic network from data
  - To find out how networks are differently associated with conditions

Individual metabolite level (1)

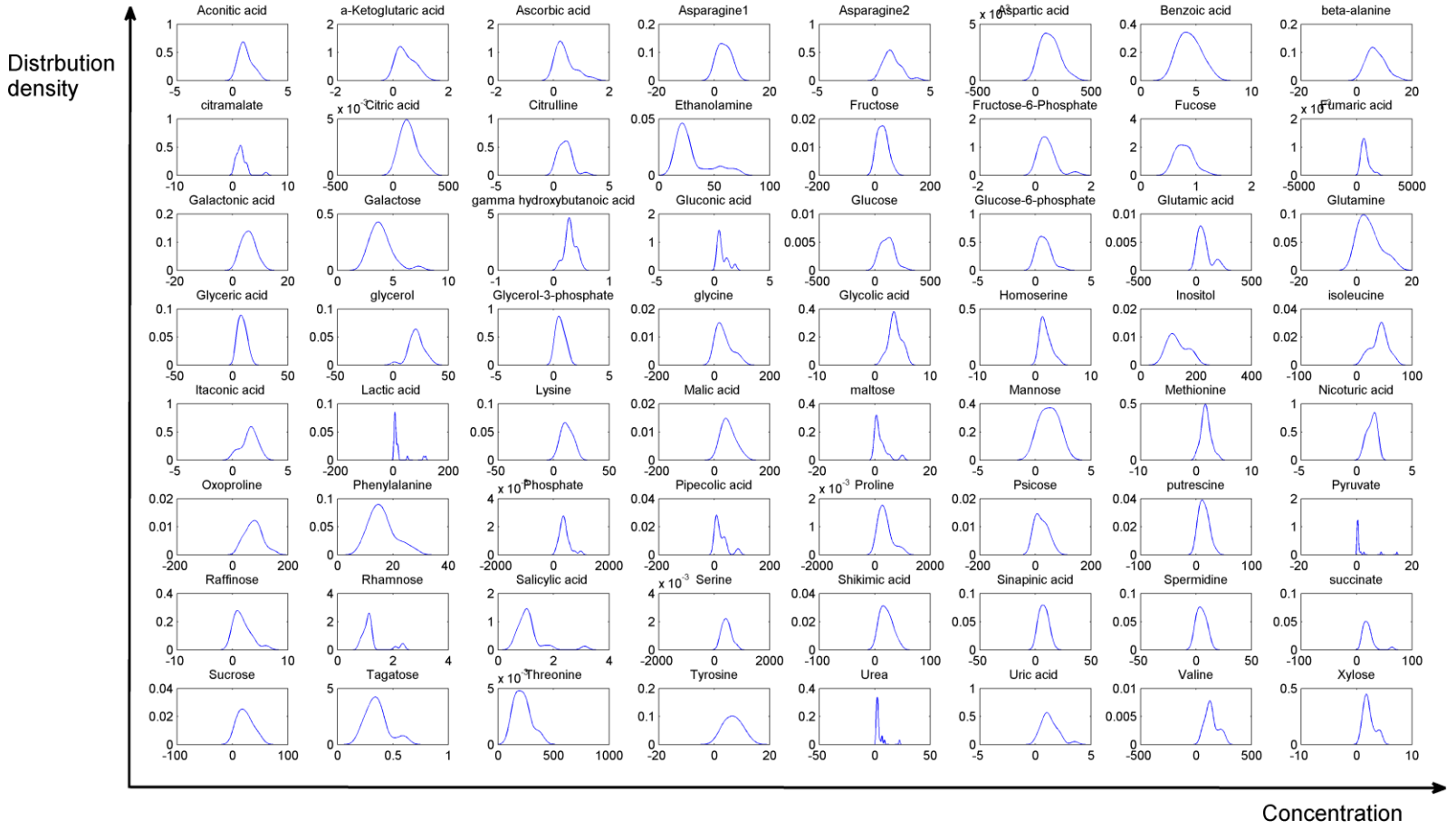
- **ANOVA** (ANalysis Of VAriance) to detect difference among conditions

	Condition 1				Condition 2				Condition 3				...	Condition n			
	R1	R2	...	Rp	R1	R2	...	Rp	R1	R2	...	Rp		R1	R2	...	Rp
Metabolite 1	x	x	x	x	x	x	x	x	x	x	x	x	...	x	x	x	x
Metabolite 2	x	x	x	x	x	x	x	x	x	x	x	x	...	x	x	x	x
Metabolite 3	x	x	x	x	x	x	x	x	x	x	x	x	...	x	x	x	x
...	...	...	...	...	...	...	...	...	...	...	...	...	...	...	...	...	...
Metabolite m	x	x	x	x	x	x	x	x	x	x	x	x	...	x	x	x	x





A metabolomics dataset

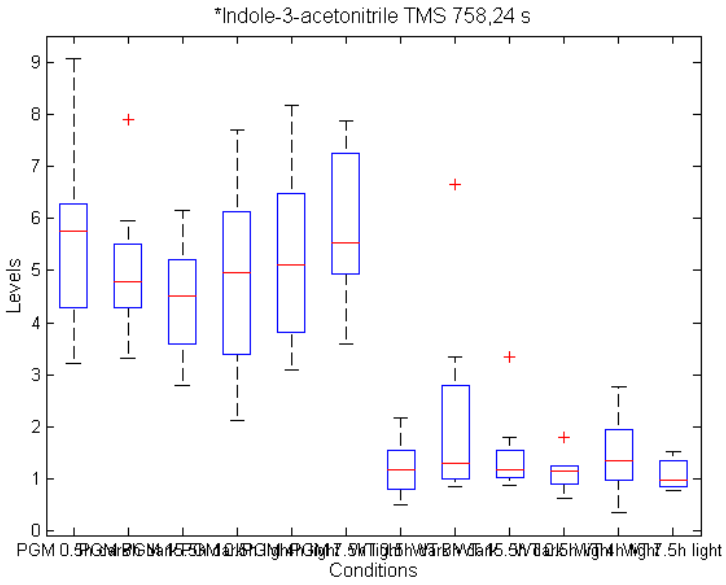


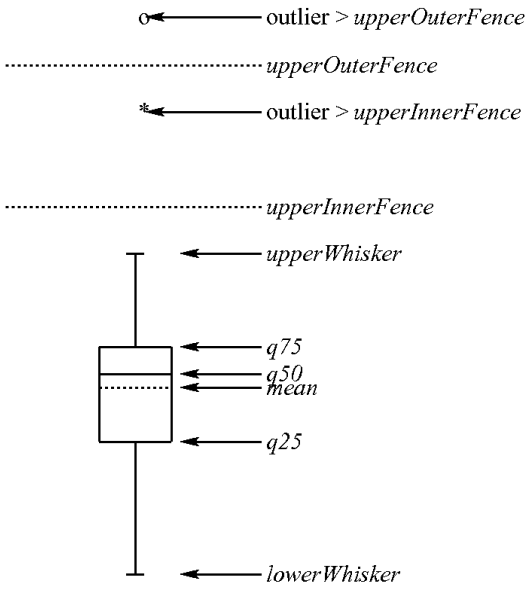
Missing values problem

- What causes missing values?
  - Because beyond machine detection ability?
  - Because software problem (e.g., deconvolution)?
  - Because randomly missed?

- Two approaches to fill missing values
  - Use half of the minimal value
  - Use distribution-based algorithm

Outliers

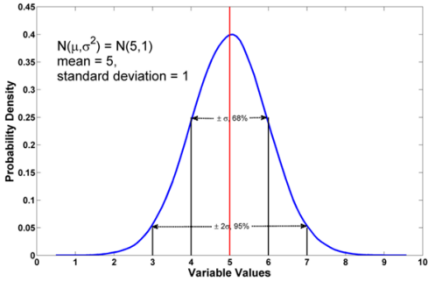




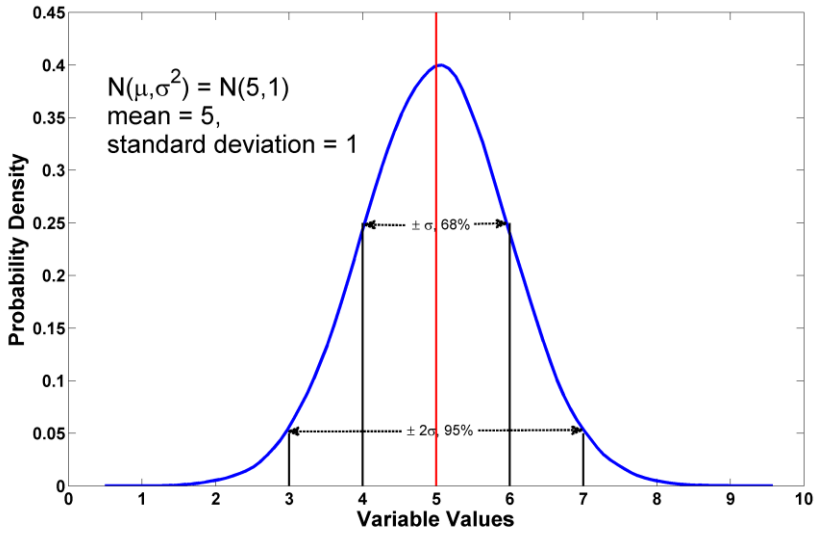
Must consider: are there biological explanations?

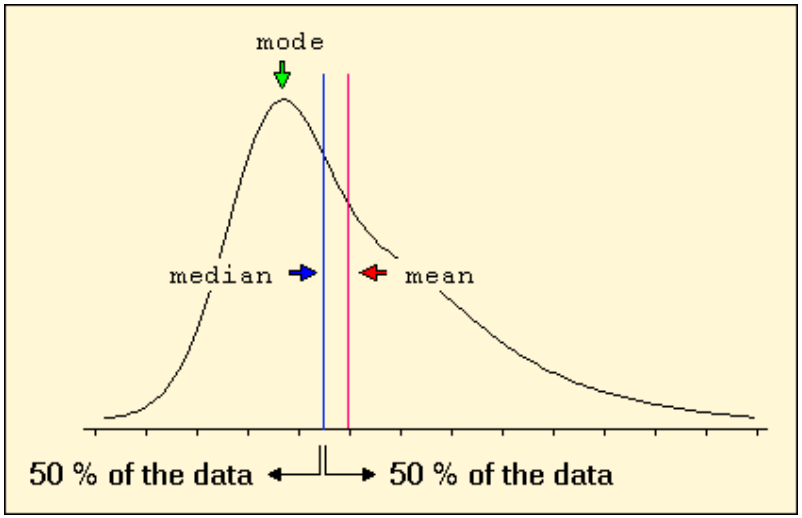
Did you know Marie Curie's story?

Adjust outliers from the proposed distribution

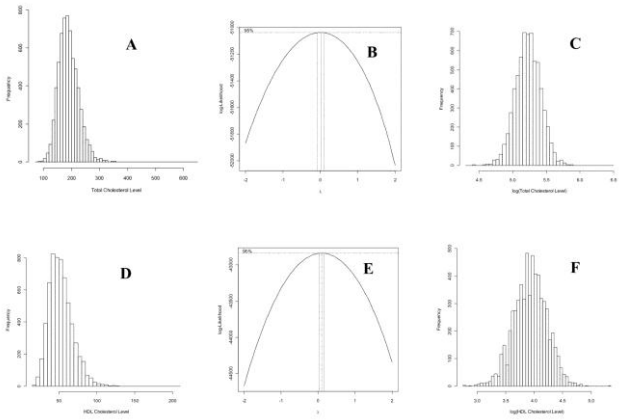


Not normally distributed data

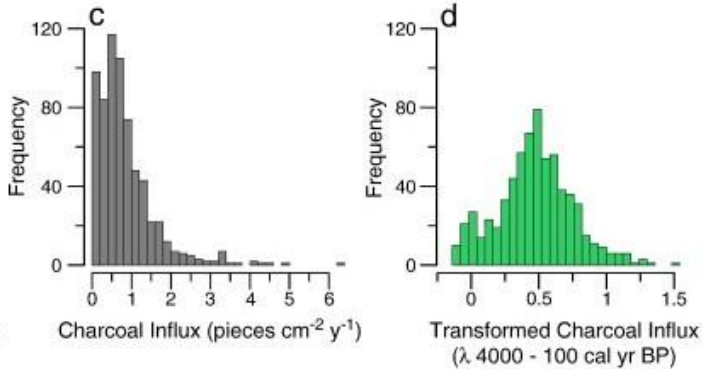




Log transformation



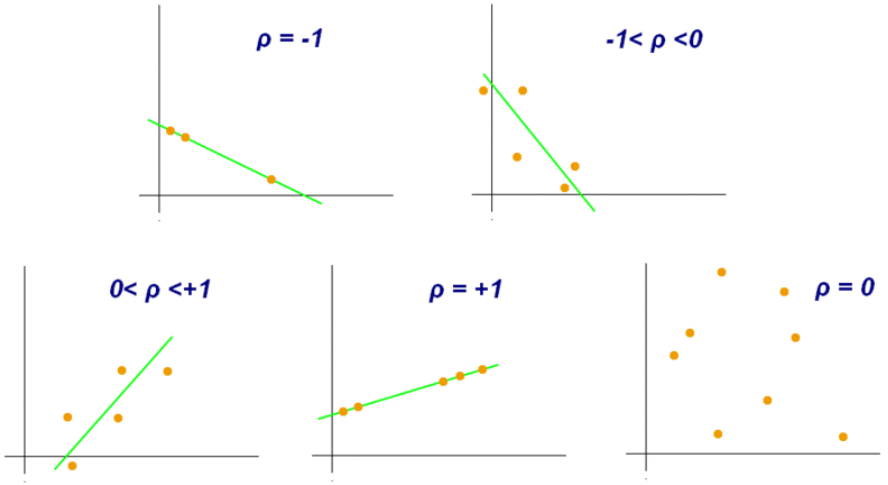
Box-cox transformation



Individual metabolite level (2)

- **Correlation coefficient** characterizes the similarity between two metabolites
  - Use Pearson's for normally distributed data
  - Use Spearman's for not normally distributed, existing outliers data.

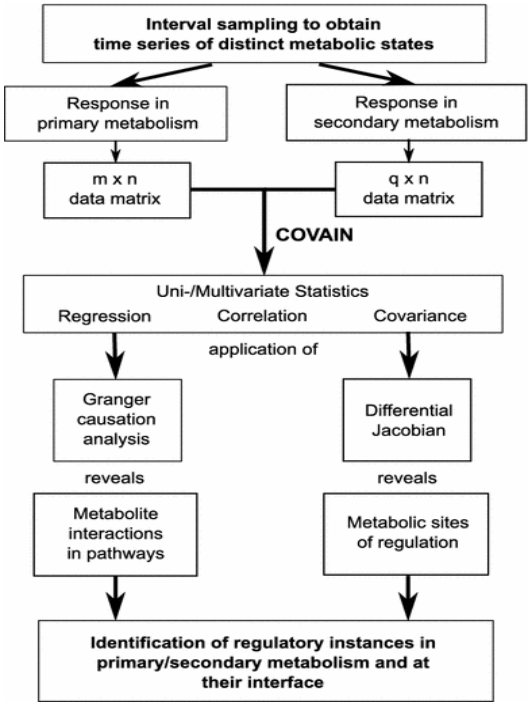
Common outputs of correlation calculation are:
1) Correlation coefficients [0-1]
2) p-value to indict significance



Individual metabolite level (3)

- **Granger causality** identifies the causation between two metabolites in time course data

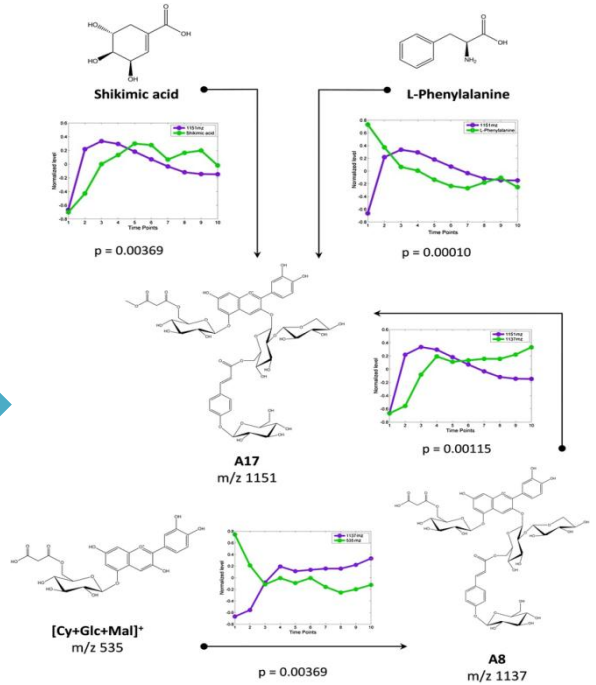
Hannes Doerfler, David Lyon*, Xiaoliang Sun*, …, Wolfram Weckwerth, Metabolomics, 2012*



$$X(t) = \sum_{i=1}^{d} C_{X,i}\, X(t-i) + \sum_{i=1}^{d} C_{XY,i}\, Y(t-i) + R_X(t)$$

$$Y(t) = \sum_{i=1}^{d} C_{YX,i}\, X(t-i) + \sum_{i=1}^{d} C_{Y,i}\, Y(t-i) + R_Y(t)$$

Granger causality in integrated GC–MS and LC–MS metabolomics data reveals the interface of primary and secondary metabolism

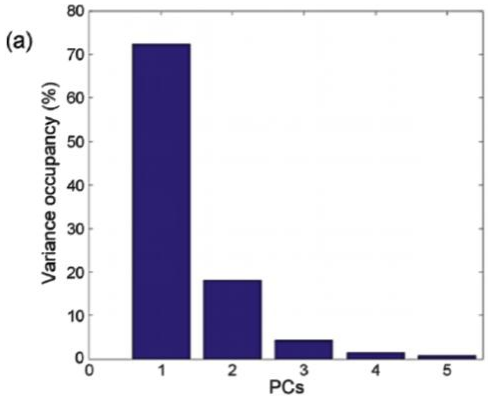


Metabolite group level (1)

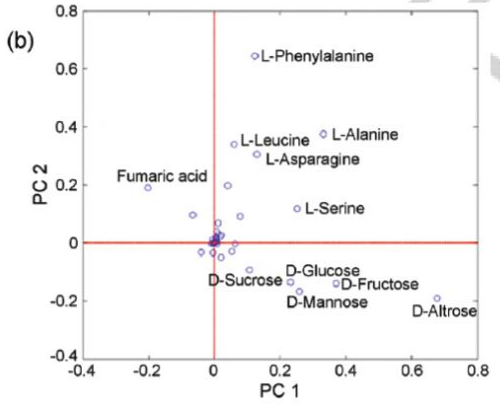
- **Principal component analysis** (PCA) distinguishes phenotypes and finds most influencing metabolites

[Loading, Coordinate, Variance] = Singular Value Decomposition(data)

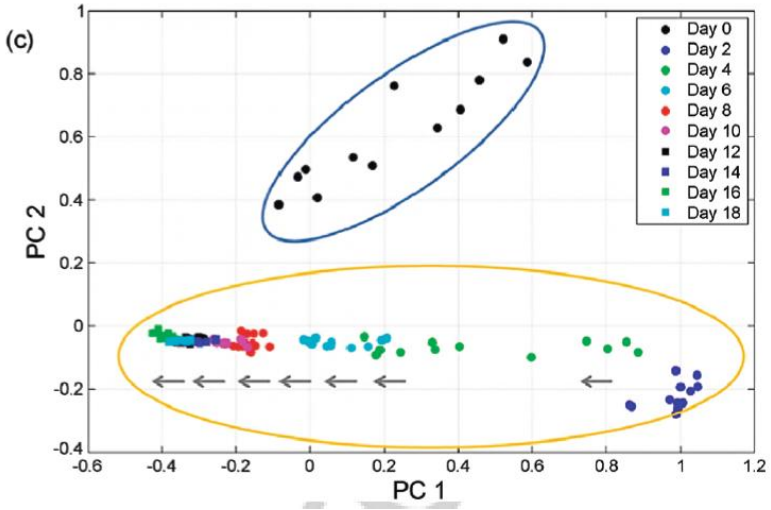
A. Variance occupancy



B. Loading plot



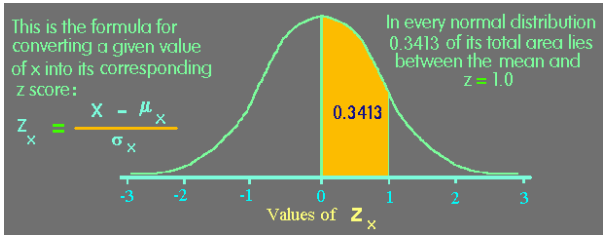
C. Coordinate plot



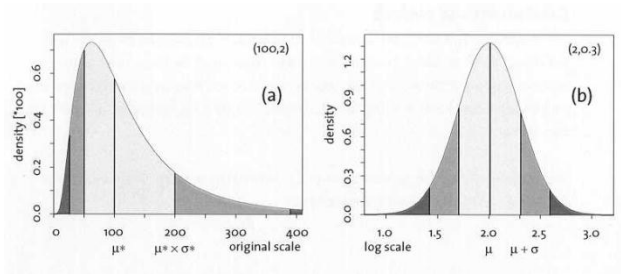
Data Scaling?

- Scaling is necessary if aiming to compare relative change

Metabolite	Initial Concentration [a.u.]	Fluctuation amplitude (%)	Changed ratio (fold)	Loading of PC1	Differential loading of PC1 (log2 ratio)
S_1	1	10	10	0.0022	**5.95**
S_2	10	10	5	0.0119	**1.64**
S_3	100	10	2	0.0268	**0.08**
S_4	**1000**	**10**	**0 (no change)**	**0.9996**	**~ 0**

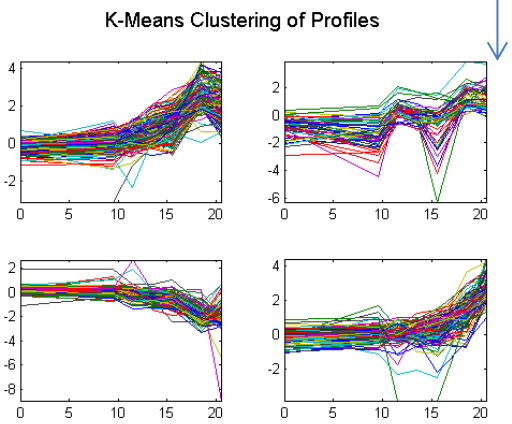


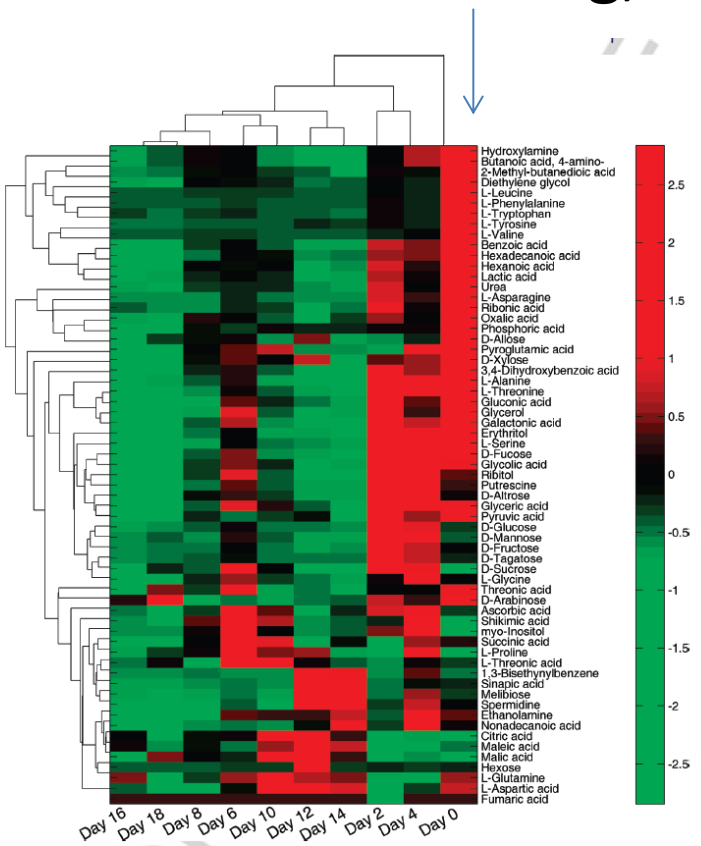
- Scaling methods:
  - Linear: z-score, range [0,1]
  - Nonlinear: log



Metabolite group level (2)

- **Clustering** classifies data into groups
  - Common methods: hierarchical clustering, k-means clustering
  - Regression
  - Based on graph





Metabolite group level (3)

- **Regression** identifies association between metabolites and conditions, and is especially useful for orthogonal experimental design

[Species]	[Tr.1]	[Tr.2]	[Tr.3]	[Tr.3]
Barke	UVA+	UVB+	PAR+	N+
Barke	UVA-	UVB-	PAR+	N+
Barke	UVA-	UVB+	PAR-	N-
Barke	UVA+	UVB+	PAR+	N-
Barke	UVA+	UVB+	PAR-	N+
Barke	UVA-	UVB-	PAR-	N-
Barke	UVA-	UVB-	PAR-	N+
Barke	UVA-	UVB-	PAR+	N-
Barke	UVA+	UVB+	PAR-	N-
Barke	UVA-	UVB+	PAR-	N+
Barke	UVA-	UVB+	PAR+	N+
Barke	UVA-	UVB+	PAR+	N-
Bonus	UVA+	UVB+	PAR+	N+
Bonus	UVA-	UVB+	PAR+	N+
Bonus	UVA-	UVB-	PAR-	N-
Bonus	UVA-	UVB-	PAR-	N+
Bonus	UVA-	UVB+	PAR+	N-
Bonus	UVA-	UVB+	PAR-	N+
Bonus	UVA-	UVB-	PAR+	N-
Bonus	UVA-	UVB+	PAR-	N-
Bonus	UVA+	UVB+	PAR-	N+
Bonus	UVA-	UVB-	PAR+	N+
Bonus	UVA+	UVB+	PAR+	N-
Bonus	UVA+	UVB+	PAR-	N-

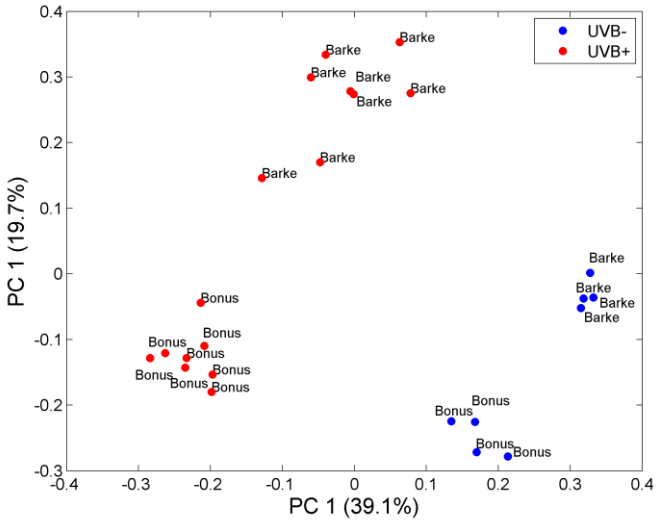
Discretize treatments: 1 for -, 2 for +

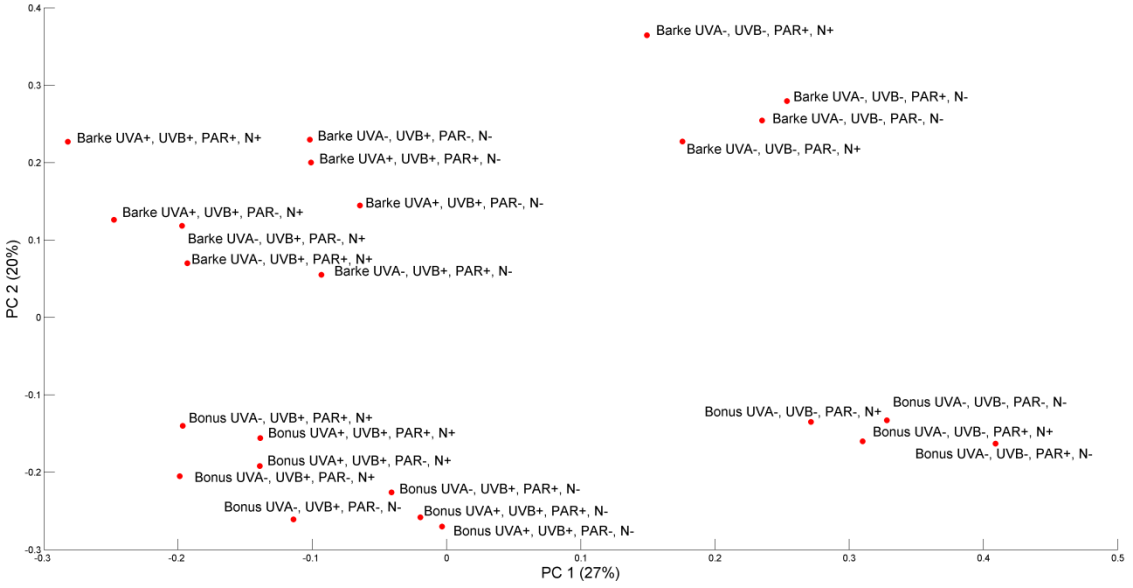
Regression equation:

m/z feature level = a0 + a1*Species + a2*Tr.1 + a3*Tr.2 + a4*Tr.3 + a5*Tr.4 + a6*Species*Tr.1 + a7*Species*Tr.2 + a8*Species*Tr.3 + a9*Tr.4 + a10*Tr.1*Tr.2 + a11*Tr.1*Tr.3 + a12*Tr.1*Tr.4 + a13*Tr.2*Tr.3 + a14*Tr.2*Tr.4 +a15*Tr.3*Tr.4 + error

Stepwise regression eliminates insignificant associations

Regression results vs. whole data





Influences of data preprocessing to statistical analysis results

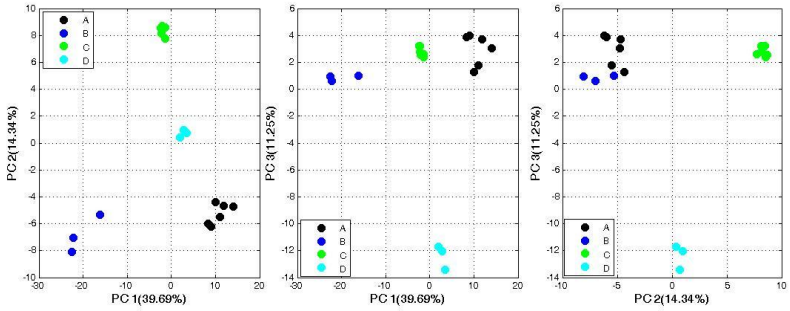
Statistical method	Metric	Missing value imputation & Outlier adjustment			Transformation & Scaling (column-wise)			
		Minimal	Mean	Distribution	Power	Log	Z-score	Range
ANOVA (t statistics)	Mean & Variance	●●	●	●	●	●	●	●
Correlation (Pearson's)	Mean	●●	○	●	●	●	○	○
Correlation (Spearman's)	Ranking of Mean	●●	○	●	○	○	○	○
Granger	Mean	●●	○	●	●	●	○	○
PCA / ICA	Covariance	●●	●●	●	●	●●	●	●
Clustering (Euclidean)	Mean	●●	○	●	●	●	○	○

●● represents strong influences, ● medium influences, and ○ no influence. Here, the ANOVA is based on a normal distribution and clustering uses the Euclidean distance.

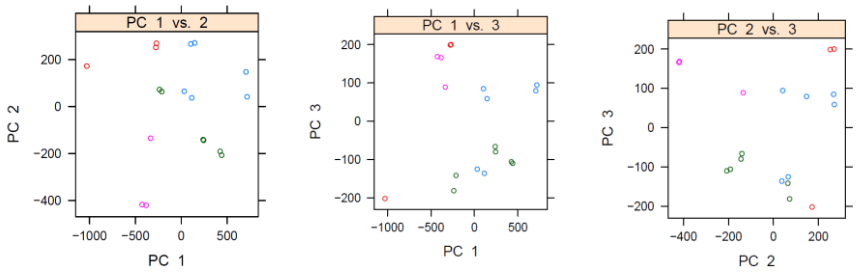
PCA analysis on a proteomics dataset

The distribution-based imputation method: using Expectation-Maximization algorithm

COVAIN results



MetaGeneAlyse results



Network level (1): database mapping

Original names

*Ribitol (5TMS) (Interner Standard)
10*13C-Sorbitol_IntSTD mz323 RT28.55
10_Aldohexose (Glucose) (1MeOx) (5TMS) BP
10_Aldohexose (Glucose) (1MeOx) (5TMS) MP
10_Sucrose (8TMS)
Aconitic acid, cis- (3TMS)?
Alanine (3TMS)
Alanine-, beta (3TMS)
Ascorbic acid 4TMS
Ascorbic acid; Dehydroascorbic acid dimer BP
Ascorbic acid; Dehydroascorbic acid dimer MP
Aspartic acid (2TMS)
Aspartic acid (3TMS)?
Citric acid (4TMS)
Cont Hexadecanoic acid
Cont Octadecanoic acid (1TMS)
Eryhronic acid (4TMS) (or Threonic acid?)
Ethanolamine (3TMS)

Reduced names

Ribitol
Sorbitol
Glucose
Sucrose
Aconitic acid, cis-
Alanine
Alanine-, beta
Ascorbic acid
Dehydroascorbic acid
Aspartic acid
Citric acid
Hexadecanoic acid
Octadecanoic acid
Eryhronic acid
Ethanolamine

KEGG names mapping

```
Metabolite: Aconitic acid, cis-

1...Octadecanoic acid
2...Echinocystic acid
3...9-cis-Retinoic acid
4...Nicotinic acid
5...cis-Aconitic acid
6...trans-Aconitic acid
7...2-Oxosuccinamic acid
8...Oxalosuccinic acid
9...2-Aminosuccinic acid
10...Citraconic acid
11...Octanedioic acid
Which one?   5
```

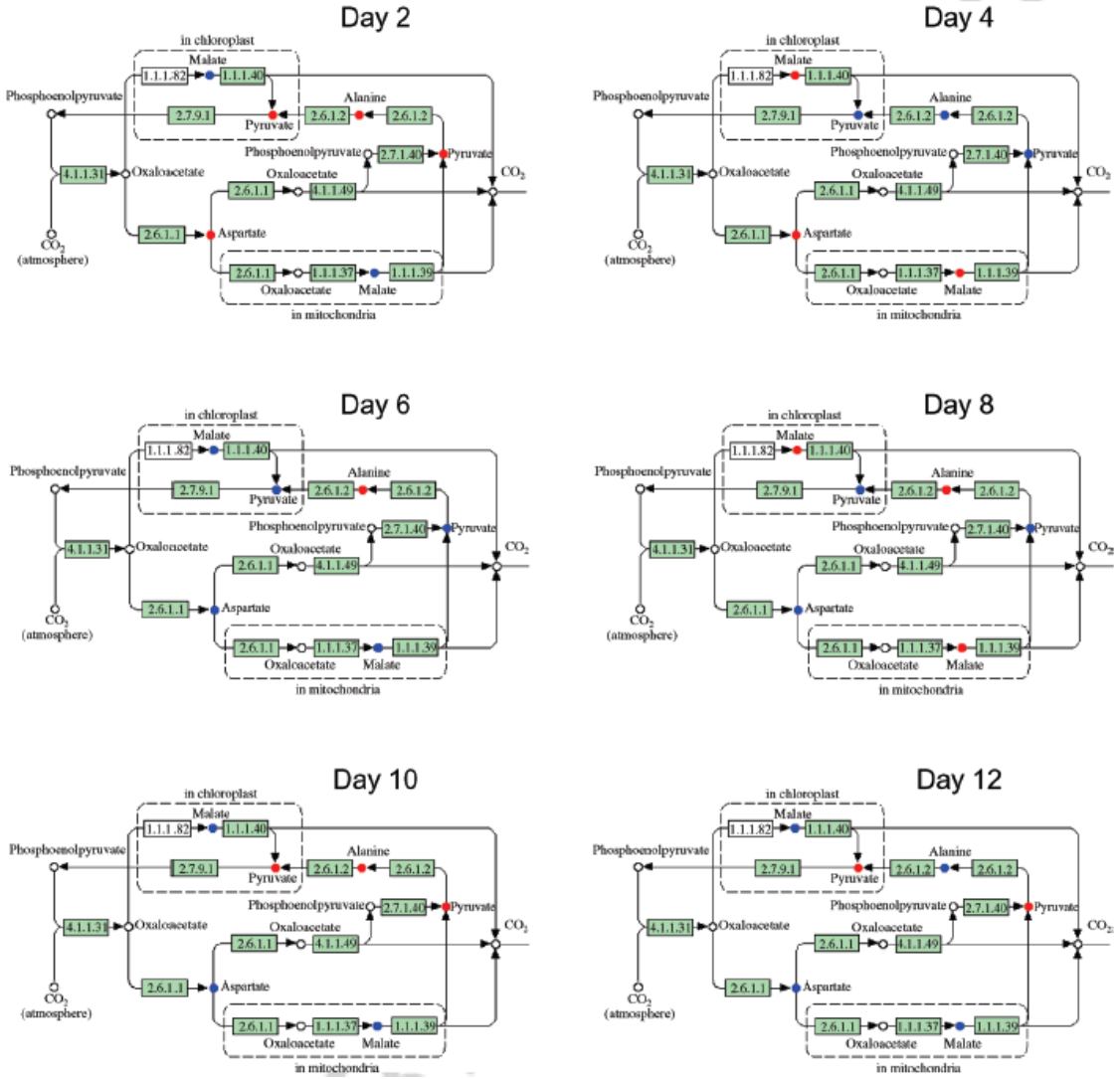
Name	KEGG Name	KEGG ID	Formula
Aconitic acid, cis-	cis-Aconitic acid	C00417	C6H6O6
Alanine	Alanine	C01401	C3H7NO2
Alanine-, beta	beta-Alanine	C00099	C3H7NO2
Ascorbic acid	Ascorbic acid	C00072	C6H8O6
Aspartic acid	Aspartic acid	C16433	C4H7NO4
Citric acid	Citric acid	C00158	C6H8O7
Dehydroascorbic acid	Dehydroascorbic acid	C05422	C6H6O6
Disaccharid	Lipid A disaccharide	C04932	C68H129N2O20P
Ethanolamine	Ethanolamine	C00189	C2H7NO
Fructose	Fructose	C02336	C6H12O6

KEGG ID/Formula Mapping

Name	KEGG Name	KEGG ID	Formula	Color	Control	3D	5D	3G	5G
Aconitic acid, cis-	cis-Aconitic acid	C00417	C6H6O6	blue	5.327144535	4.77204931	4.75735824	5.41641324	5.03003
Alanine	Alanine	C01401	C3H7NO2	blue	5.768923669	5.31855579	5.42984623	6.0864222	6.705485
Alanine-, beta	beta-Alanine	C00099	C3H7NO2	blue	5.18895195	4.8628535	4.94017365	5.35239817	5.535512
Ascorbic acid	Ascorbic acid	C00072	C6H8O6	blue	5.355646654	5.33068737	5.54944019	5.45579956	5.716773
Aspartic acid	Aspartic acid	C16433	C4H7NO4	red	5.835499666	6.07170371	6.73390743	6.14484235	5.97393
Citric acid	Citric acid	C00158	C6H8O7	blue	7.470201132	7.29674425	7.25328882	7.5181206	7.303258
Dehydroascorbic acid	Dehydroascorbic acid	C05422	C6H6O6	blue	6.450444957	6.42382375	6.8047005	6.58694452	6.465055
Disaccharid	Lipid A disaccharide	C04932	C68H129N2O20P	blue	4.423030873	4.05857819	4.32859405	4.4992042	4.078352
Ethanolamine	Ethanolamine	C00189	C2H7NO	blue	7.53389473	7.46042205	7.51635012	7.57852577	7.445617
Fructose	Fructose	C02336	C6H12O6	blue	6.568186542	6.41723942	6.97633475	6.54992489	6.457776

User's data integration

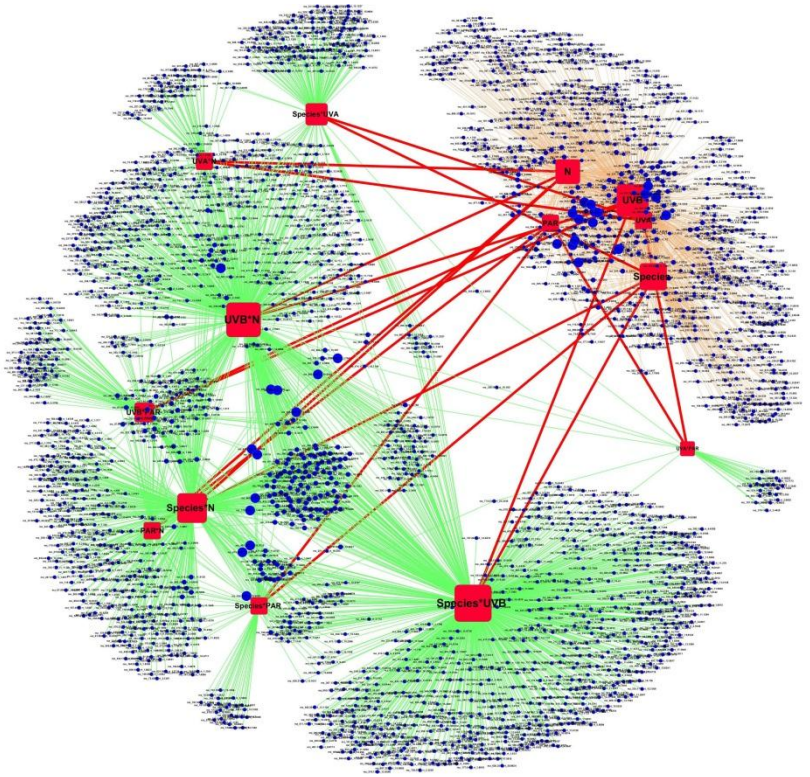
An example. Red: up- / Blue: down- regulated



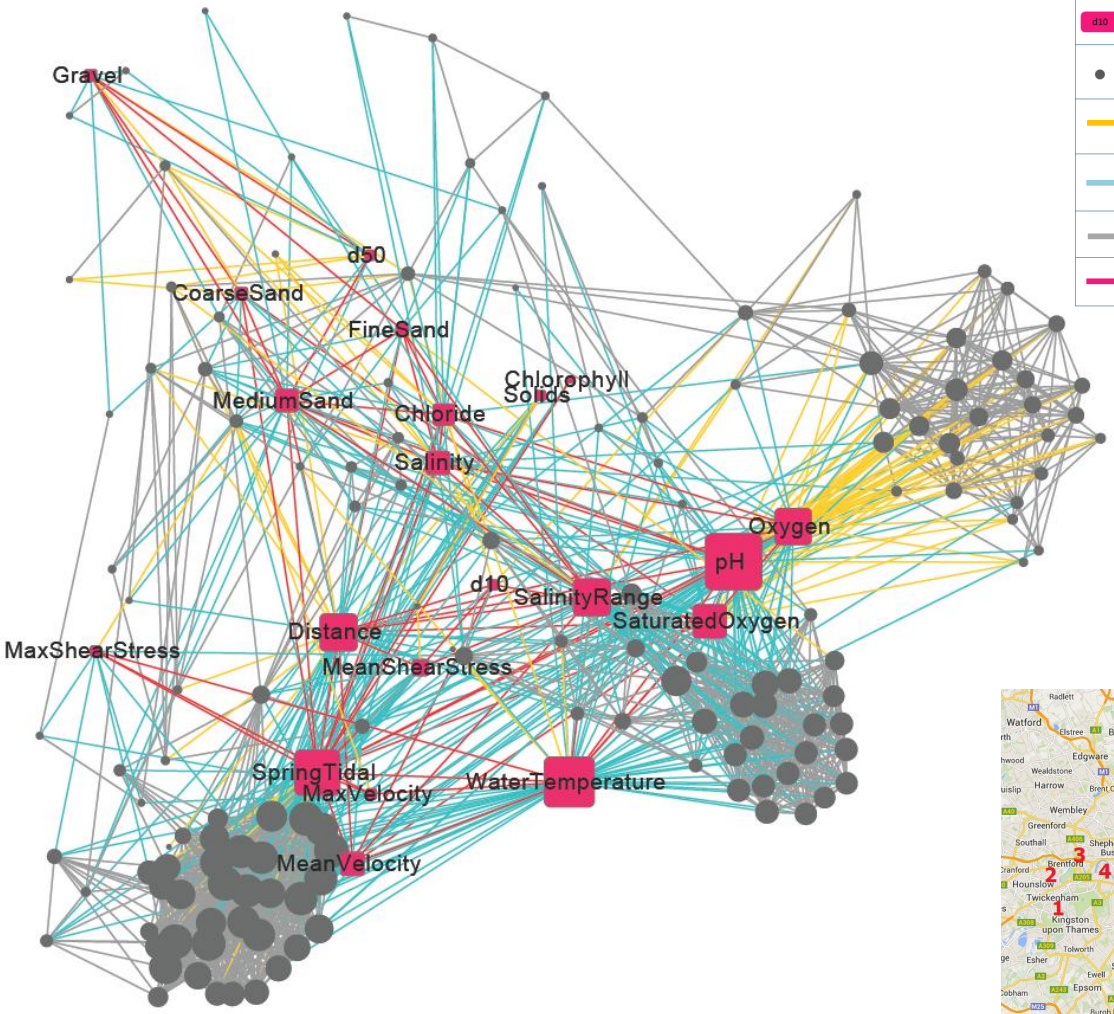
Pathway enrichment analysis can be done afterward.

Network level (2): inferring from data

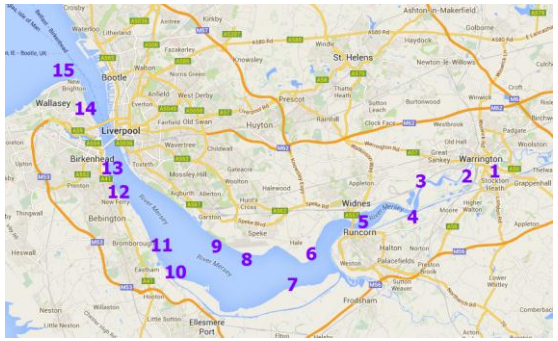
- Based on:
  - Correlation coefficients
  - Granger causality
  - Regression

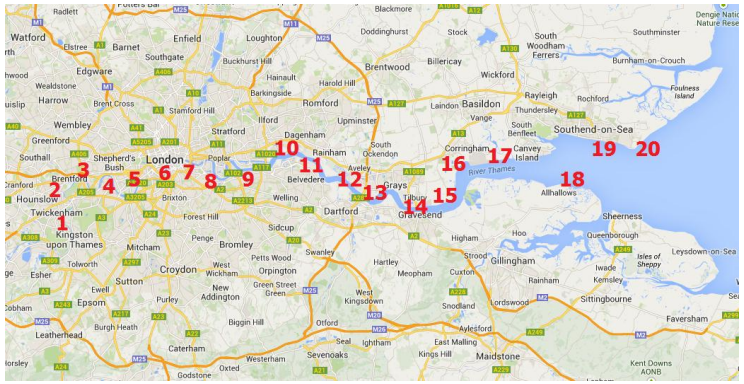


An ecological network

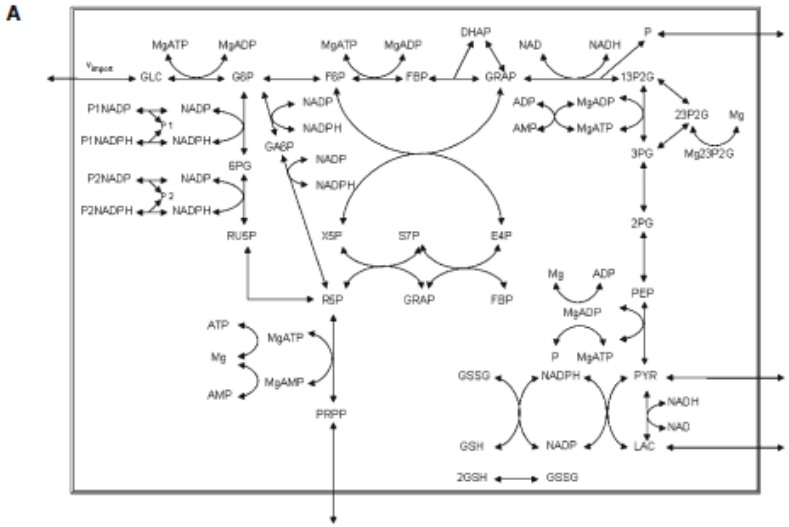


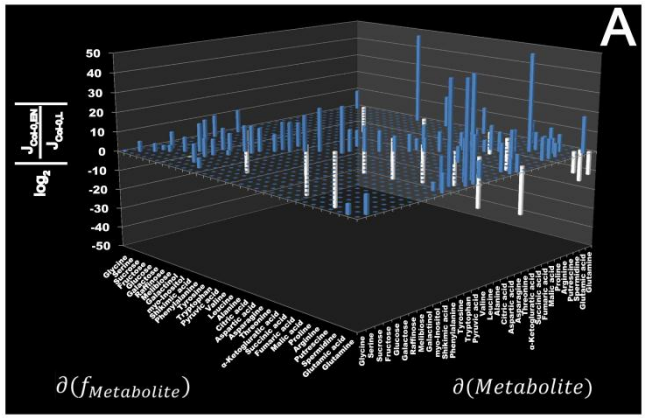
d10 pH	Environmental factors. Size proportional to degree.
• •	OTUs. Size proportional to degree.
	Associations between OTU and one environmental factor.
	Associations between OTU and multiple environmental factors.
	Associations between OTUs.
	Associations between environmental factors.

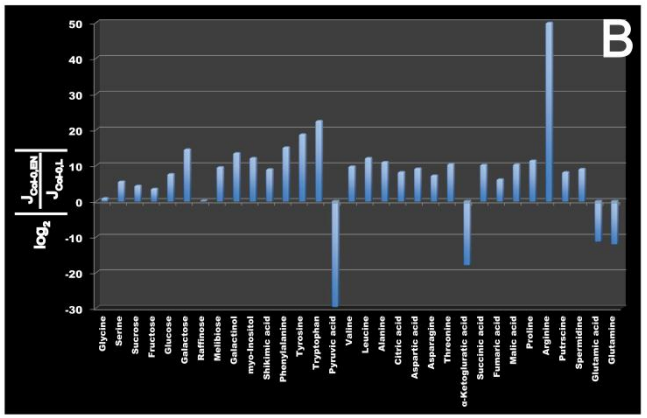




Network level (3): differential Jacobian identifies dynamical changes in metabolic network

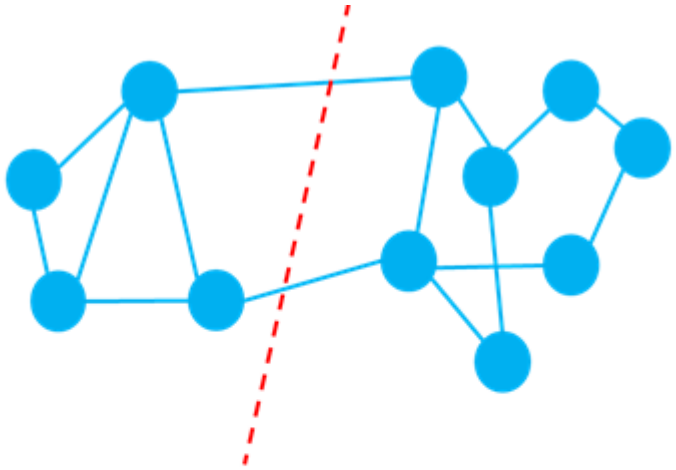


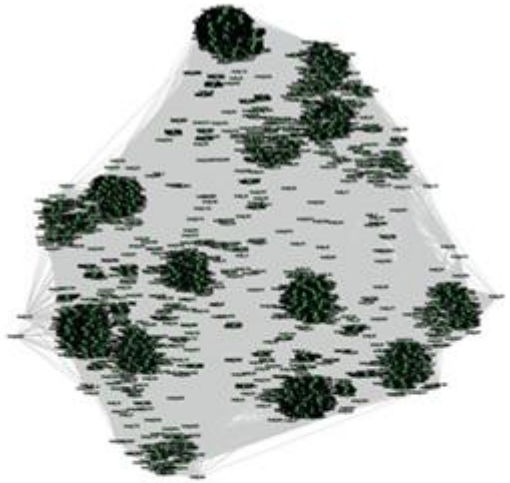




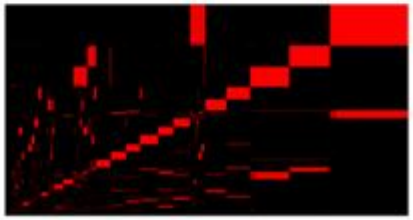
Sun X, Weckwerth W, Metabolomics, 2012
Nagele T, Mair A, Sun X, et al, PLoS One, 2014

Network level (4): spectral clustering partitions network into separate modules

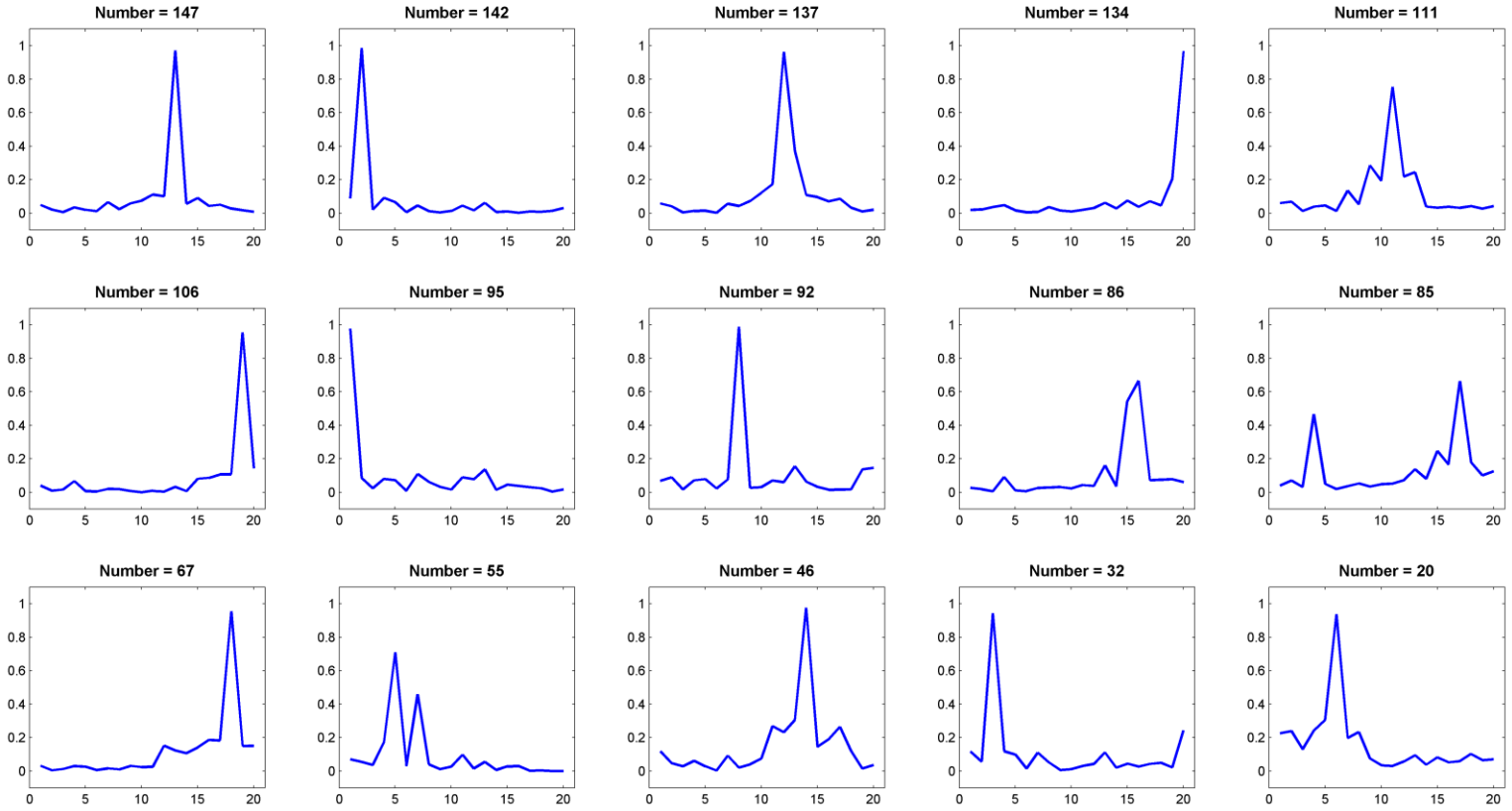




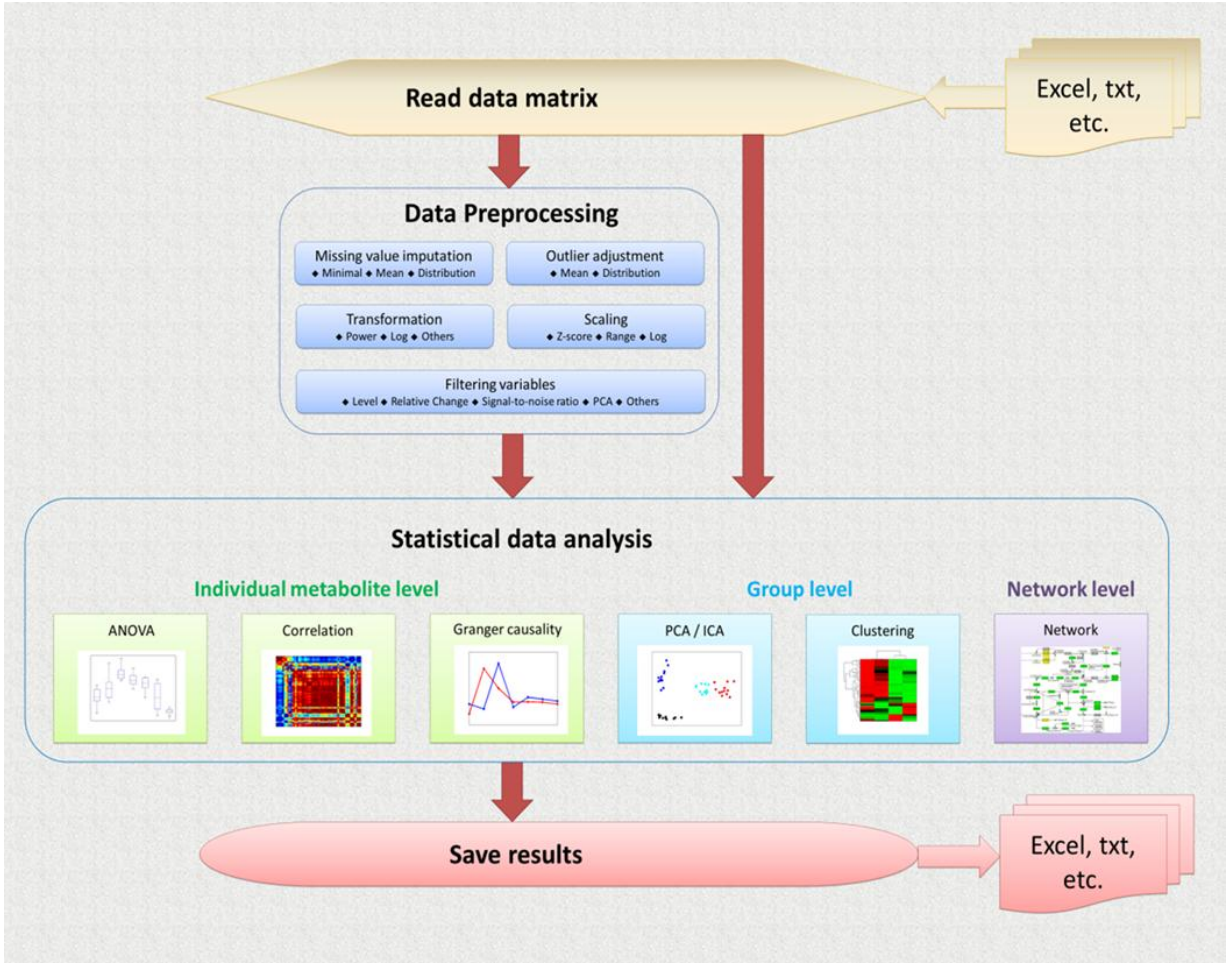
The network is constructed from regression coefficients and is partitioned into a few modules by spectral clustering.



Clustering variables from network



Analyzing metabolomics data by COVAIN



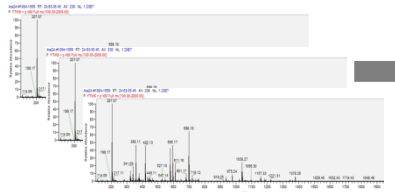
Sun X & Weckwerth W, Metabolomics, 2012
Doerfler H*, Sun X*, et al, PLoS One, 2014

A short illustration on analyzing cold stress GC-MS data

Most methods presented in this talk are included in the COVAIN software.

mzGroupAnalyzer - Predicting Pathways and Novel Chemical Structures from Untargeted High-Throughput Metabolomics Data

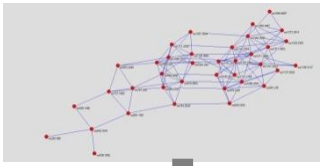
Read m/z data and reaction rules



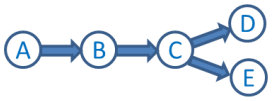
m/z	Intensity	Relative	Theo. Mass	Delta (ppm)	Composition
448.11559	25978.6	0.86	448.11578	-0.41	C11 H22 O14 N5
			448.11527	0.72	C25 H20 O8
			448.11526	0.73	C24 H14 O3 N7
449.10783	188211.2	6.2	449.10783	-0.01	C20 H15 O6 N7
			449.10784	-0.02	C21 H21 O11
594.18977	9240.6	0.3	594.18977	-0.01	C36 H26 O5 N4
			594.19028	-0.86	C22 H28 O11 N9
			594.19028	-0.87	C23 H34 O16 N2
594.35642	8594.1	0.28	594.35644	-0.04	C37 H46 O3 N4
			594.35695	-0.89	C23 H48 O9 N9
			594.35696	-0.9	C24 H54 O14 N2
595.14474	71219.1	2.35	595.14462	0.21	C30 H27 O13
			595.14461	0.22	C29 H21 O8 N7
			595.14512	-0.64	C16 H29 O19 N5

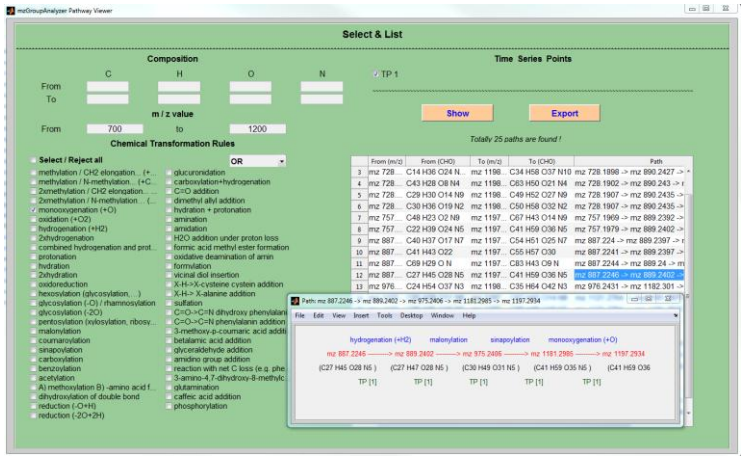
Reaction step / chemical modification	net C change	net H change	net O change	net N change
methylation / CH2 elongation... (+CH2)	1	2	0	0
methylation / N-methylation... (+CH3)	1	3	0	0
2xmethylation / CH2 elongation... (+ 2xCH2)	2	4	0	0
2xmethylation / N-methylation... (+2xCH3)	2	6	0	0
monooxygenation (+O)	0	0	1	0
oxidation (+O2)	0	0	2	0
hydrogenation (+H2)	0	2	0	0
2xhydrogenation	0	4	0	0
X-H->X-cysteine cystein addition	3	5	2	1
X-H-> X-alanine addition	3	5	2	1
sulfation	0	0	3	0
combined hydrogenation and protonation	0	3	0	0
protonation	0	1	0	0
hydration	0	2	1	0
2xhydration	0	4	2	0
oxidoreduction	0	-2	1	0
hexosylation (glycosylation,...)	6	10	5	0
glycosylation (-O) / rhamnosylation	6	10	4	0

mzGroupAnalyzer + Pathway Viewer

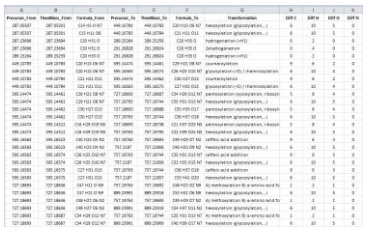


A --> B X
A --> B --> C X
A --> B --> C --> D ✓
A --> B --> C --> E ✓





Outputs



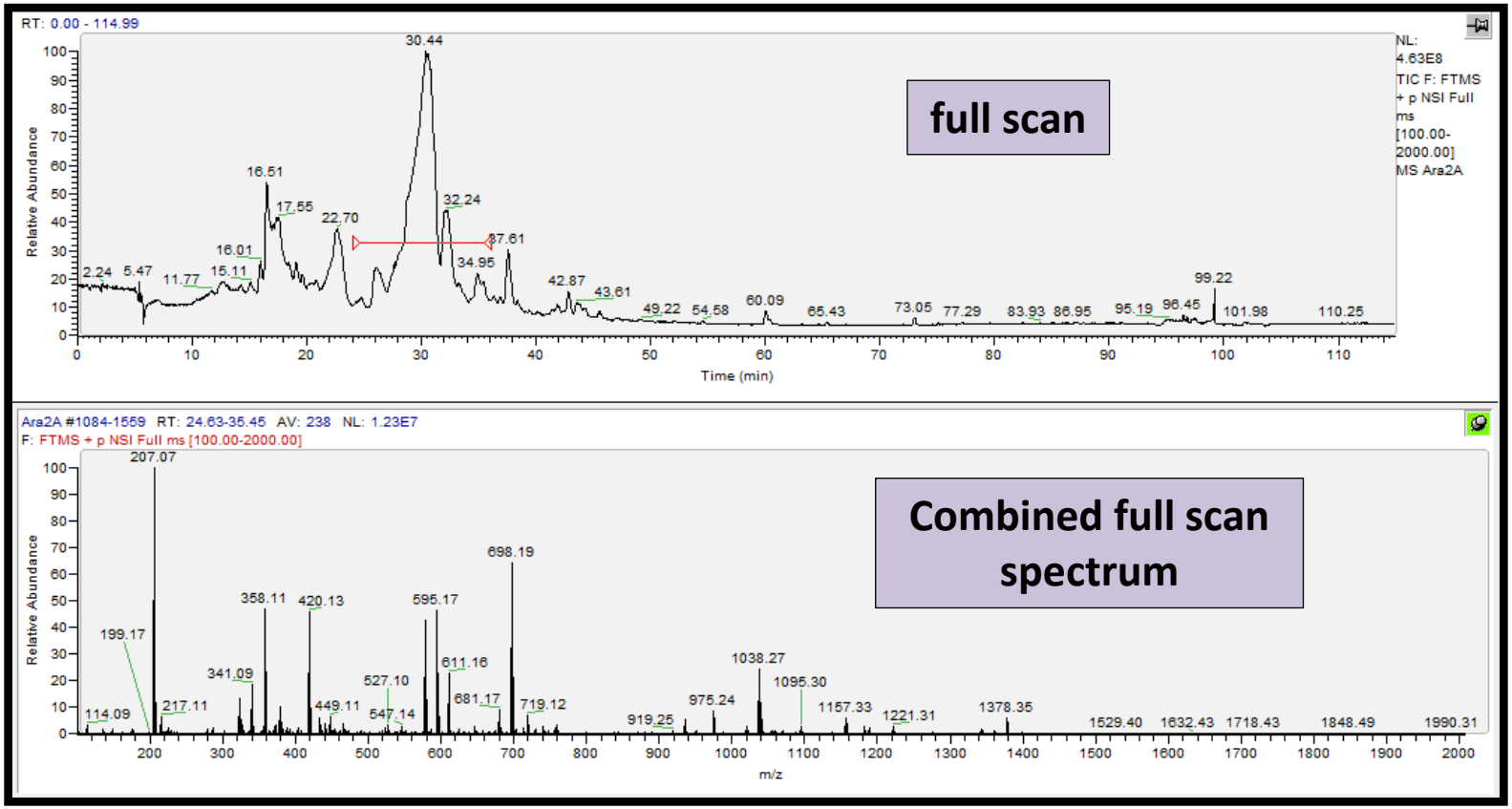
Count	Name	Diff C	Diff H	Diff O	Diff N
43	hexosylation (glycosylation,...)	6	10	5	0
38	malonylation	3	2	3	0
38	A) methoxylation B) α-amino acid	1	2	1	0
36	methylation / CH2 elongation... (	1	2	0	0
28	monooxygenation (+O)	0	0	1	0
24	coumaroylation	9	6	2	0
21	vicinal diol insertion	2	4	2	0
17	2xhydrogenation	0	4	0	0
16	sinapoylation	11	10	4	0
10	hydrogenation (+H2)	0	2	0	0
10	carboxylation-hydrogenation	0	2	2	0
9	pentosylation (xylosylation, ribo	5	8	4	0
6	3-methoxy-p-coumaric acid addit	10	8	3	0
3	caffeic acid addition	9	6	3	0
2	glycosylation (-O) / rhamnosylati	6	10	4	0
2	amination	0	1	0	1
0	methylation / N-methylation... (+	1	3	0	0
0	2xmethylation / CH2 elongation...	2	4	0	0

Count	Formular	Diff C	Diff H	Diff O	Diff N
171	-C12H10O8N2	-12	-91	-89	-2
171	C12H10O8N2	12	91	89	2
148	-C14H9O8N99	-14	-93	-89	-99
148	C14H9O8N99	14	93	89	99
137	-C13H97O93N95	-13	-97	-93	-95
137	C13H97O93N95	13	97	93	95
136	-CH6ON93	-1	-6	-4	-93
136	CH6ON93	1	6	4	93
133	-C27H85O79N	-27	-85	-79	-1
133	C27H85O79N	27	85	79	1
82	-C4H98O97N98	-4	-98	-97	-98
82	C4H98O97N98	4	98	97	98
75	-C8H98O95N97	-8	-98	-95	-97
75	C8H98O95N97	8	98	95	97
73	-C28H91O83N94	-28	-91	-83	-94
73	-C13H87O85N6	-13	-87	-85	-6
73	-C2H2N97	-2	-2	0	-97

```
data: [489x5 double]
dataAll: {489x5 cell}
finalCorrectPath: {3329x1 cell}
finalTimedPath: {3329x1 cell}
mmNet: [489x489 logical]
mmNetS: [489x489 double]
mmNetPattern: [489x489 double]
mzList: [489x1 double]
mzName: {489x1 cell}
ruleList: {54x1 cell}
x_final: [3329x1 double]
y_final: [3329x1 double]
allPath: [3329x1 struct]
compReal: {'C' 'H' 'O' 'N' 'S'}
```

1. **Select the desired m/z range in full scan for a combined full-scan spectrum (whole chromatogram might be too data-heavy – divide in more work steps in this case)**

- Choose full scan in chromatogram window
- „Pin“ spectrum window below and select TIC range with left mouse hold down



2. Set parameters for your data in „spectrum list“ view

- In spectrum window: right click → view → spectrum list
- In „View“ header → open „Info bar“ → „Elemental composition“
- Parameters might depend on sample type, acquisition method, etc.
- Tick and untick „Elemental composition“ to refresh parameter settings

Parameter settings (red windows in next slide) e.g.:

- elements C,H,O,N,S
- Intensity 0.5-100 %
- 5 ppm, 1-2 or lower for high res instruments
- 5 formulae
- …

Thermo Xcalibur Qual Browser - [Ara2A.RAW]

File Edit View Display Grid Actions Tools Window Help

Elemental composition

Single mass

Mass:

Max. results 10 Calculate

Idx	Formula	RDB	Delta mmu

File... List Simulate

Limits

Charge: 1

Nitrogen-Rule: Do not use

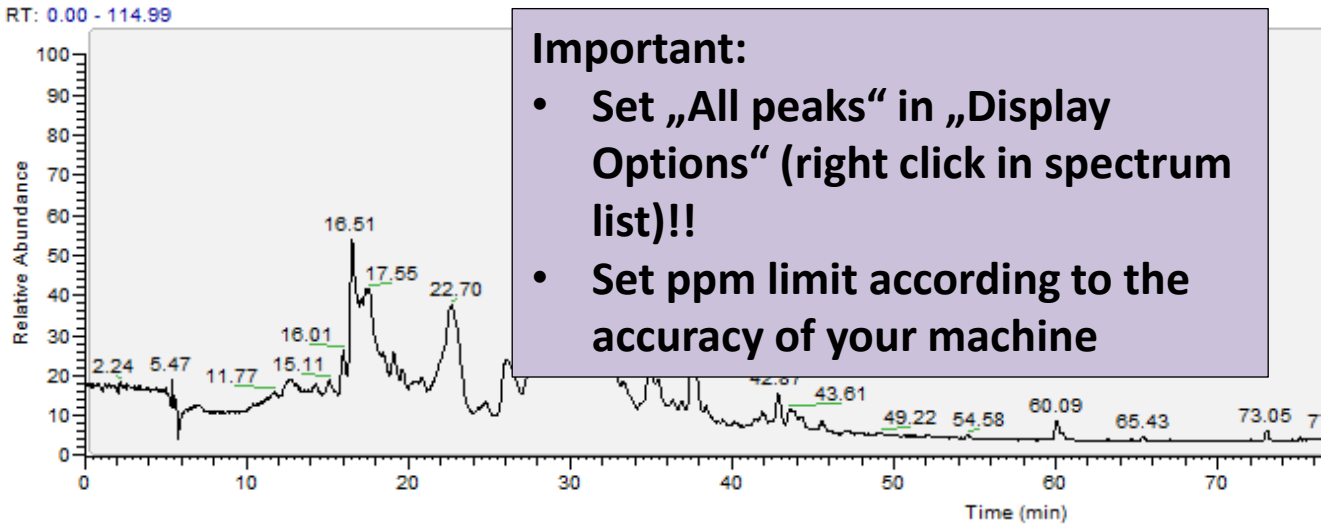
Mass tolerance 140.00 mmu

RDB equiv: -1.0-100.0

Elements in use

Isotope	Min	Max	DB eq.	Mass
14 N	0	10	0.5	14.003
16 O	0	15	0.0	15.995
12 C	0	30	1.0	12.000

C:\Users\...\LC_Daten raw\Ara2A 14.06.2011 13:49:05



Important:

- **Set „All peaks“ in „Display Options“ (right click in spectrum list)!!**
- **Set ppm limit according to the accuracy of your machine**

Ara2A#617-1723 RT: 14.03-39.23 AV: 554

F: FTMS + p NSI Full ms [100.00-2000.00]

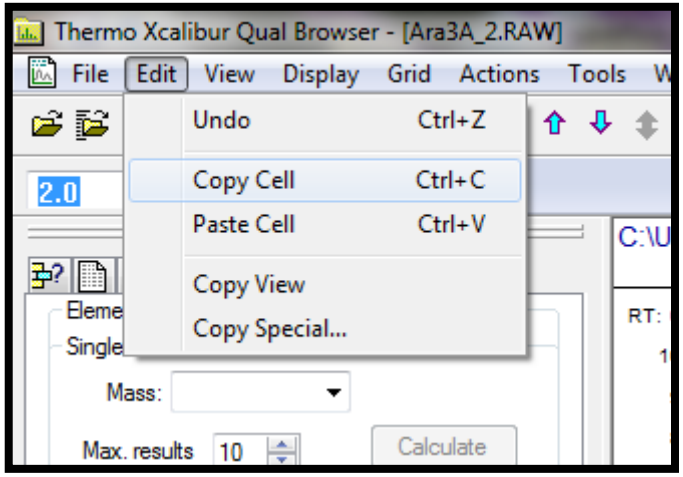
m/z	Intensity	Relative	Theo. Mass	Delta (ppm)	Composition
100.07584	181551.6	3.24	100.07569	0.15	$C_5H_{10}ON$
114.09143	403032.5	7.20	114.09134	0.09	$C_6H_{12}ON$
135.10163	261628.9	4.67	135.10157	0.06	$C_6H_{15}O_3$
136.02160	240385.4	4.29	136.02271	-1.11	$CH_4O_4N_4$
147.04411	122086.8	2.18	147.04406	0.06	$C_9H_7O_2$
149.02336	236602.6	4.23	149.02332	0.04	$C_8H_5O_3$
163.13291	142901.4	2.55	163.13287	0.04	$C_8H_{19}O_3$
177.05465	155743.0	2.78	177.05462	0.03	$C_{10}H_9O_3$
193.04954	379139.6	6.77	193.04954	0.01	$C_{10}H_9O_4$

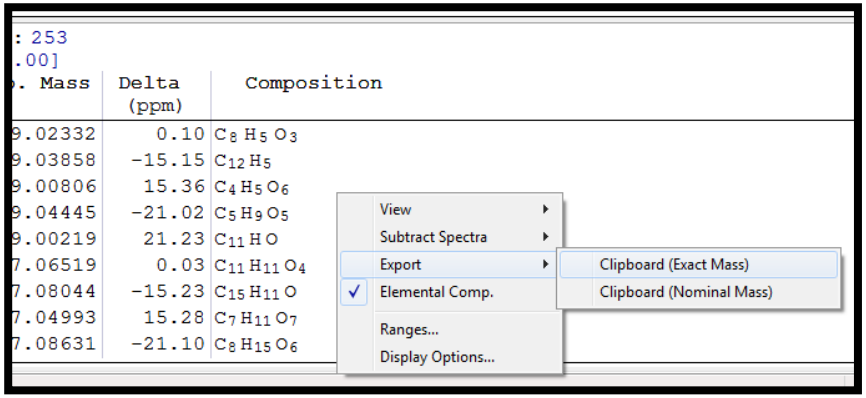
3. Export from Xcalibur to Excel

- „Copy cell“ in Xcalibur „Edit“ header (seems to be best option RAM-storage-wise)

alternatively

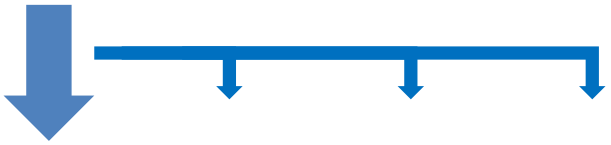
- „Export to Clipboard (Exact mass)“



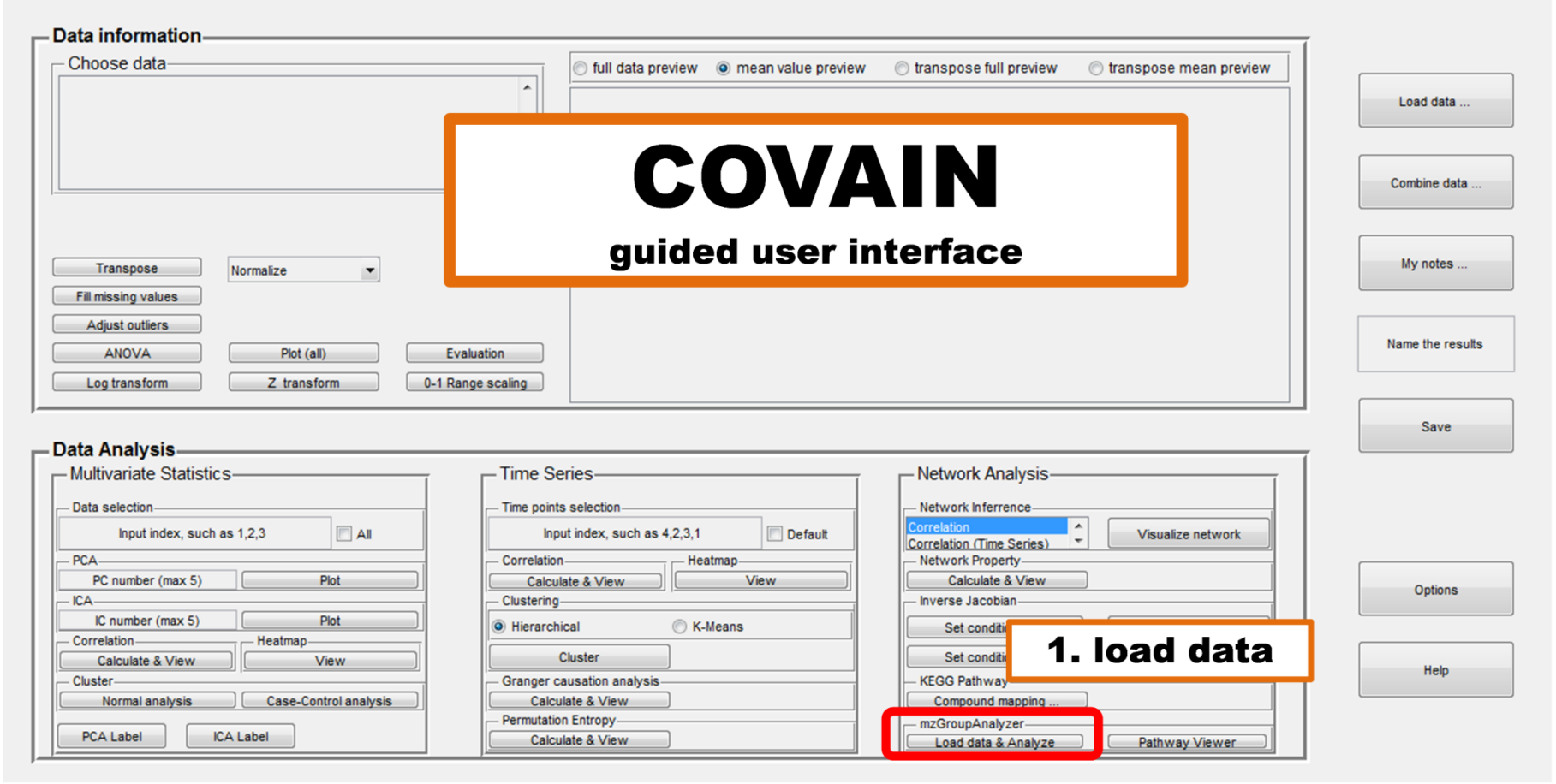


4. Prepare Excel sheet for mzGroupAnalyzer

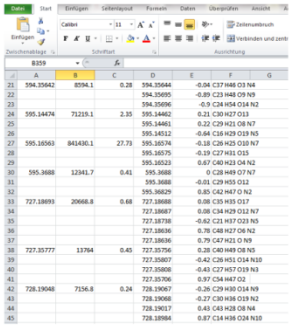
Separate the elemental composition by using „text to columns“ and „space“ as separator. You have to keep the „composition“ column, so copy/paste it into the next column and then separate it



A	B	C	D	E	F	G	H	I
m/z	Intensity	Relative	Theo. Mass	Delta (ppm)	Composition	C	H	O
149.02343	171190.6	2.04	149.02332	0.1	C8 H5 O3	C8	H5	O3
			149.03858	-15.15	C12 H5	C12	H5	
			149.00806	15.36	C4 H5 O6	C4	H5	O6
			149.04445	-21.02	C5 H9 O5	C5	H9	O5
			149.00219	21.23	C11 H O	C11	H	O
207.06521	4045081.5	48.13	207.06519	0.03	C11 H11 O4	C11	H11	O4
			207.08044	-15.23	C15 H11 O	C15	H11	O
			207.04993	15.28	C7 H11 O7	C7	H11	O7
			207.08631	-21.1	C8 H15 O6	C8	H15	O6
			207.04406	21.16	C14 H7 O2	C14	H7	O2
208.06854	471427.2	5.61	208.07301	-4.47	C11 H12 O4	C11	H12	O4
			208.05775	10.79	C7 H12 O7	C7	H12	O7

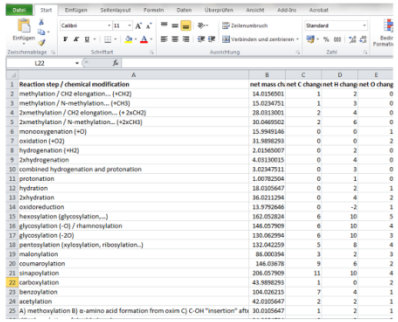


2. select data, rule file and output location



data file

\+

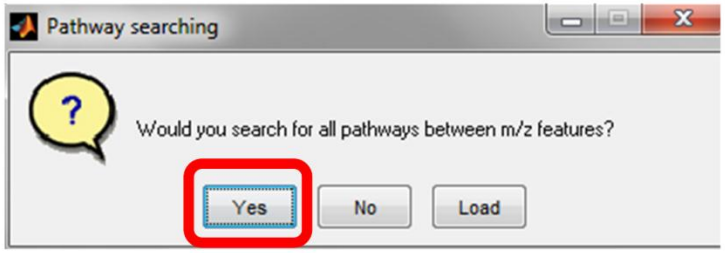


rule file

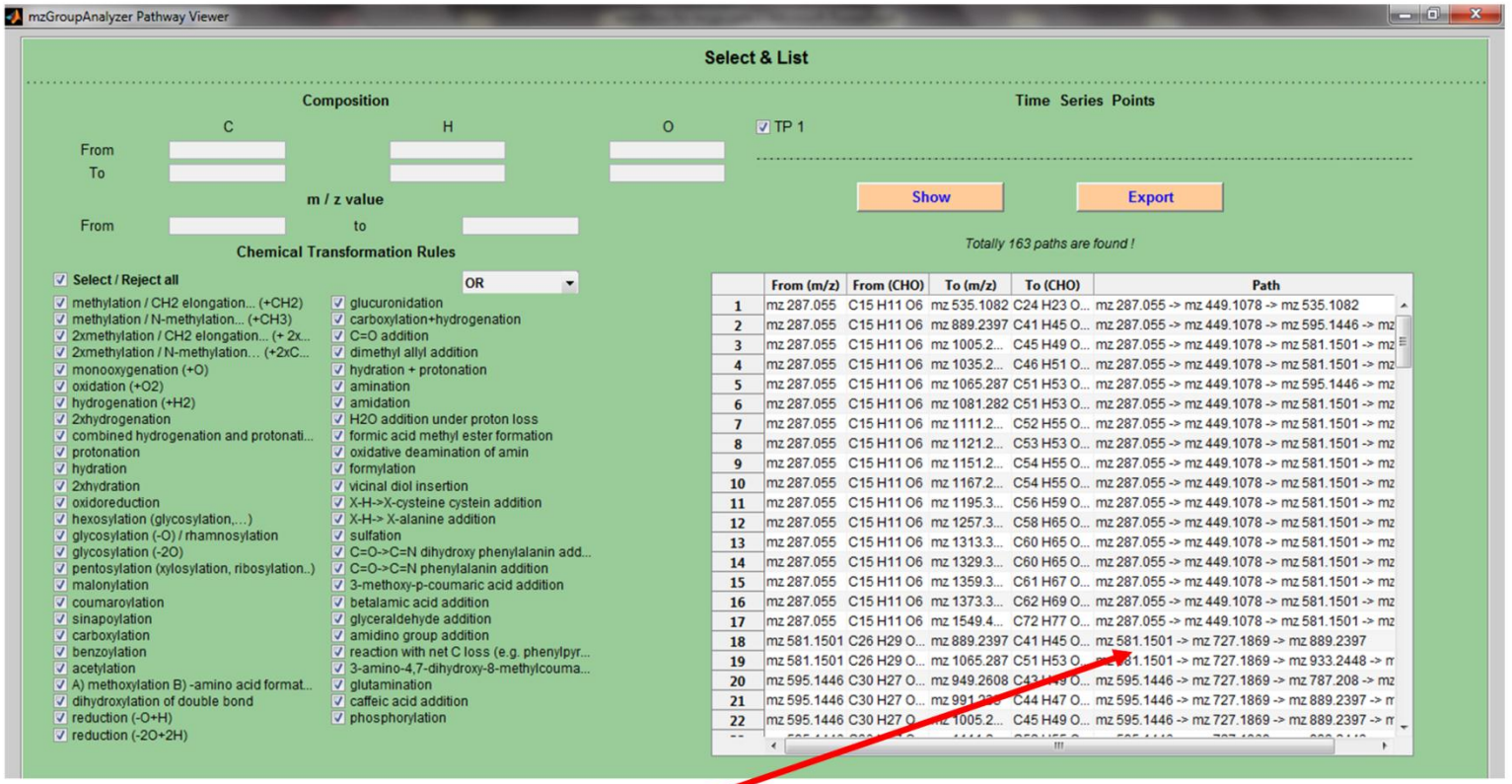
\+ output directory

→ transformation files will be created

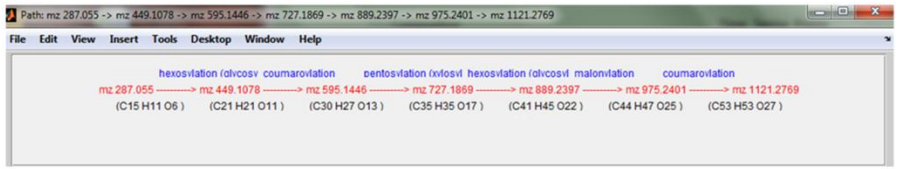
3. start *Pathway Viewer*



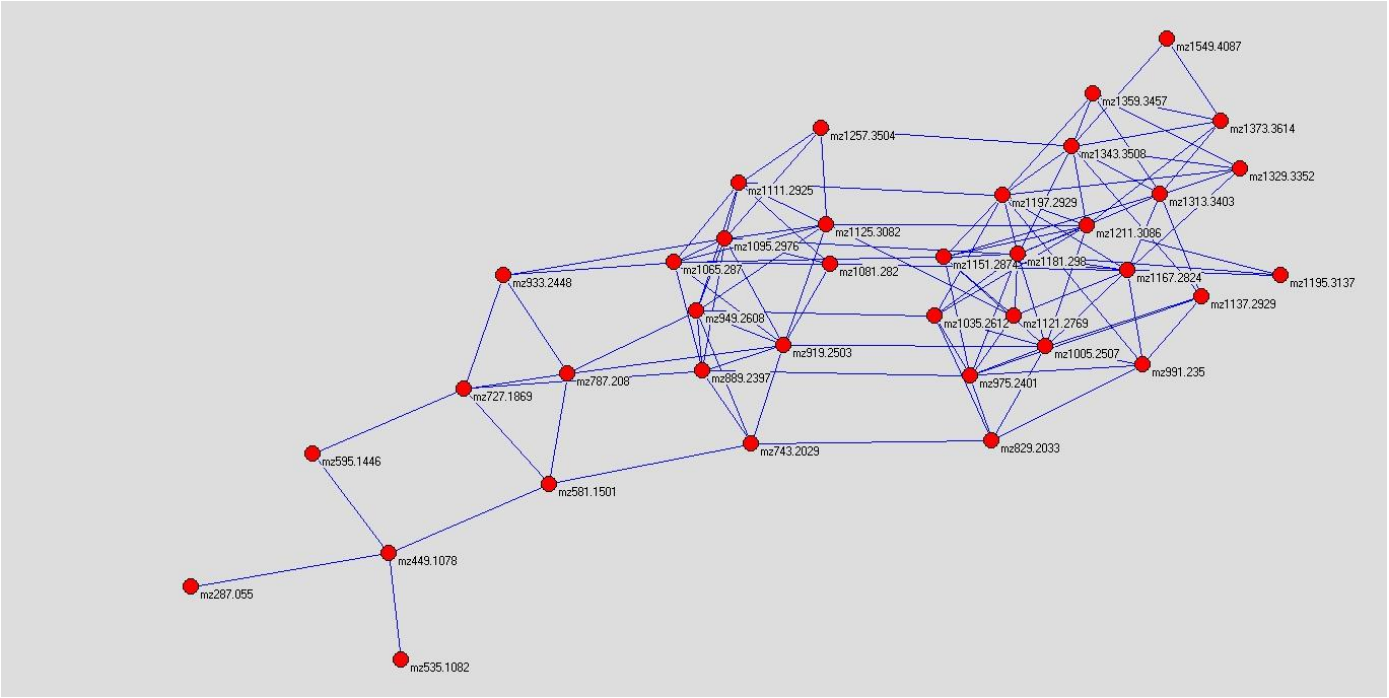
or when mzStruct.mat file has already been created before:



clicking into the „path“ header will highlight the selected path



4. view the Pajek.net file with Pajek



Finally, find the results in your specified saving folder.

mzGroup_test214	14/02/2014 13:25	85 KB
mzGroup_test214_freqListedGroups.xlsx	14/02/2014 13:25	11 KB
mzGroup_test214_freqNotListedGroups.xlsx	14/02/2014 13:25	47 KB
mzGroup_test214_mzStruct	14/02/2014 13:26	128 KB
mzGroup_test214_Pajek.net	14/02/2014 13:25	14 KB
mzGroup_test214_transformations.xlsx	14/02/2014 13:25	29 KB

Acknowledgements

- Czech Globe
  - Dr. Otmar Urban
  - Dr. Michal Oravec
  - Eva Jurkova & Zdeňka Cermanová

- University of Vienna
  - Prof. Dr. Wolfram Weckwerth
  - Hannes Doerfler

And you for your attention!

Plan the future

- I will add more advanced methods, e.g., supervised machine learning, to COVAIN
- I am willing to develop methods and software for your data
- Tutorial on COVAIN can be done in 2 hours.