



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.

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.)

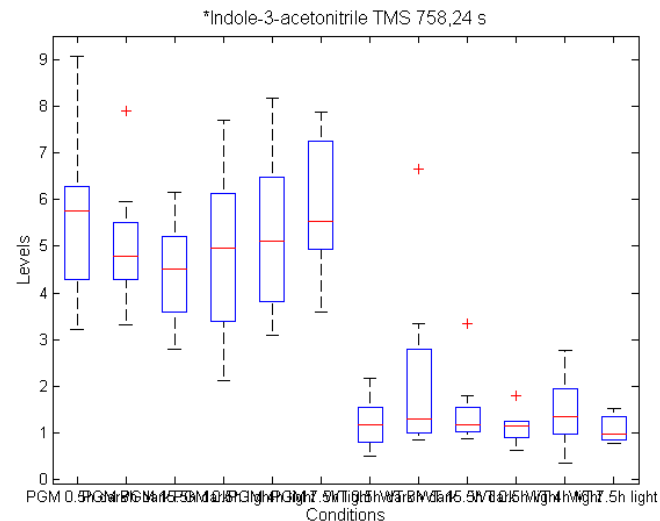
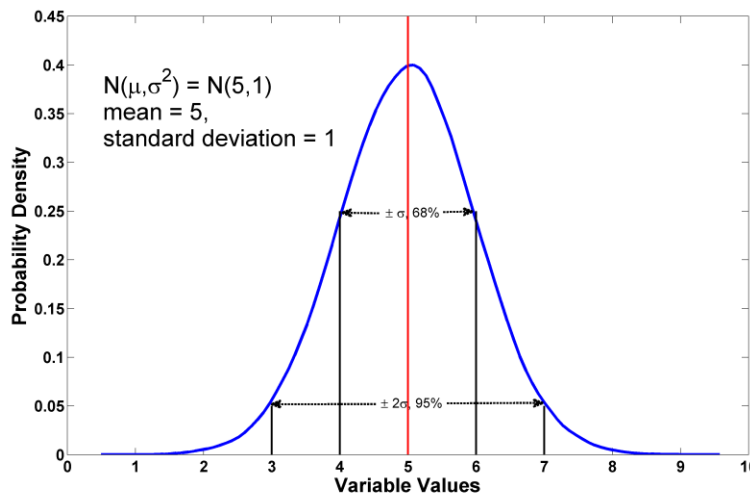
Principles and methods to analyze metabolomics data

- **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

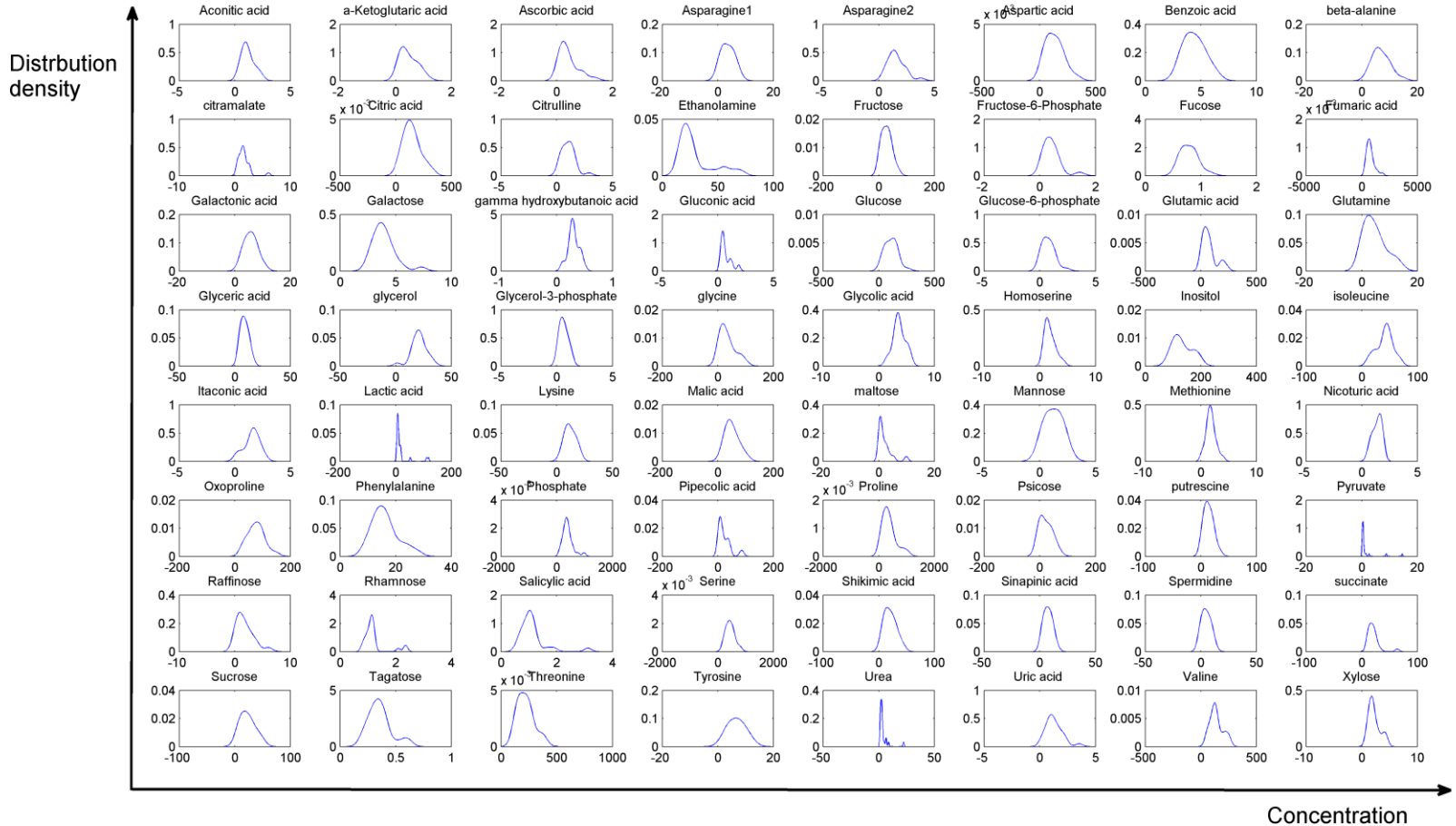
Individual metabolite level (1)

- **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



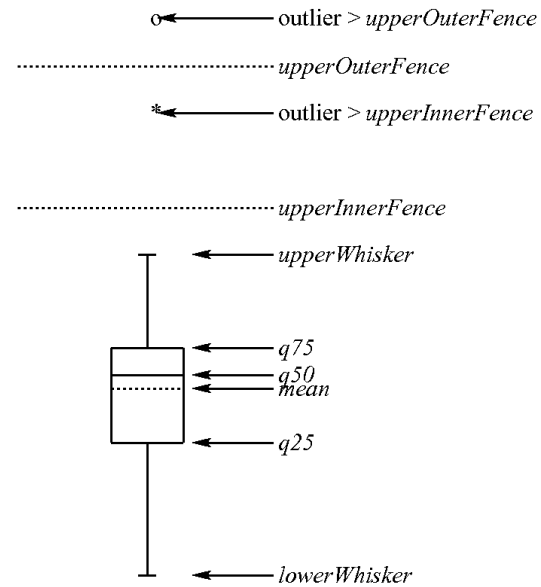
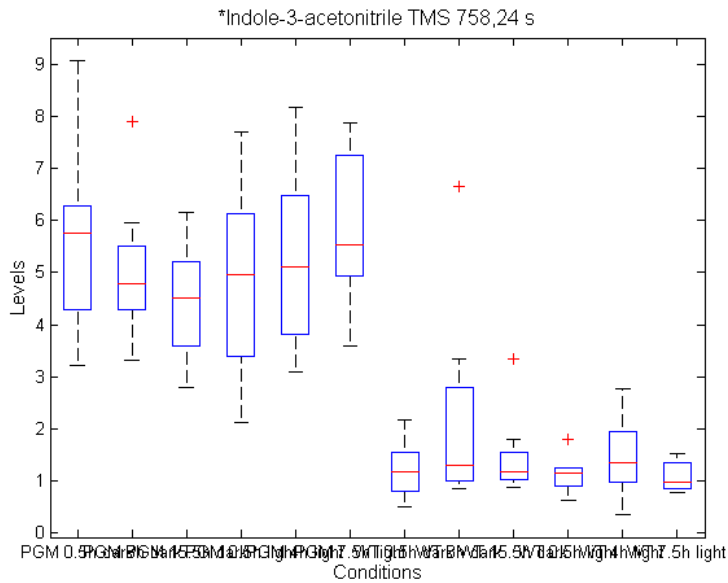
A metabolomics dataset



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

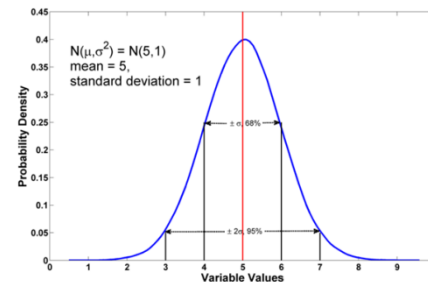


Must consider: are there biological explanations?

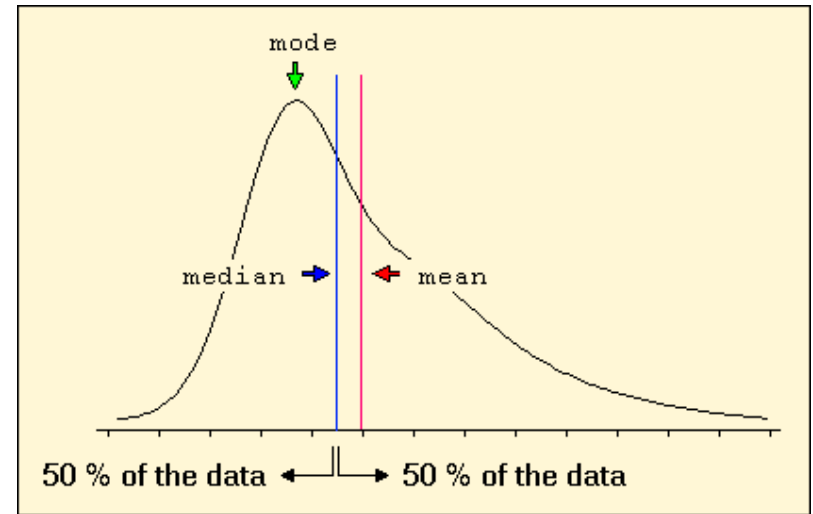
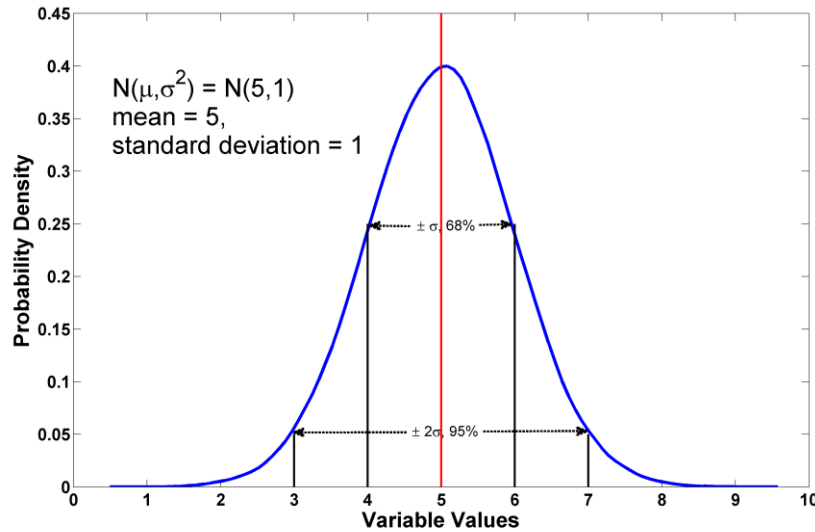
Did you know Marie Curie's story?



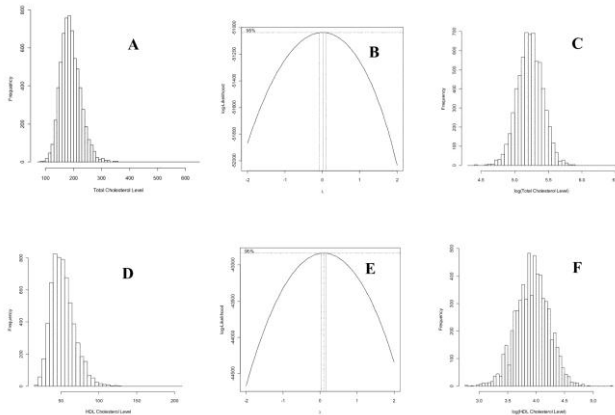
Adjust outliers from the proposed distribution



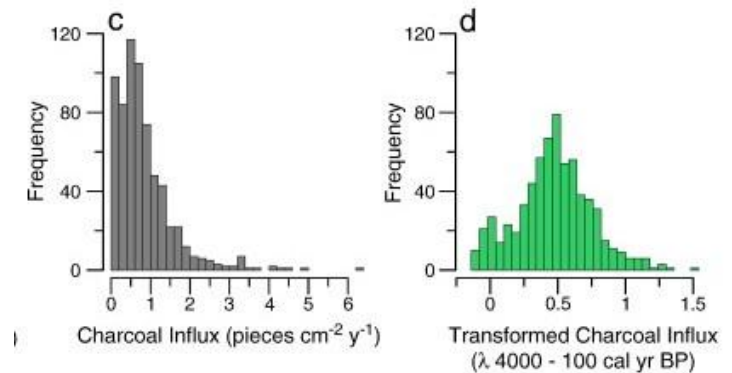
Not normally distributed data



Log transformation



Box-cox transformation

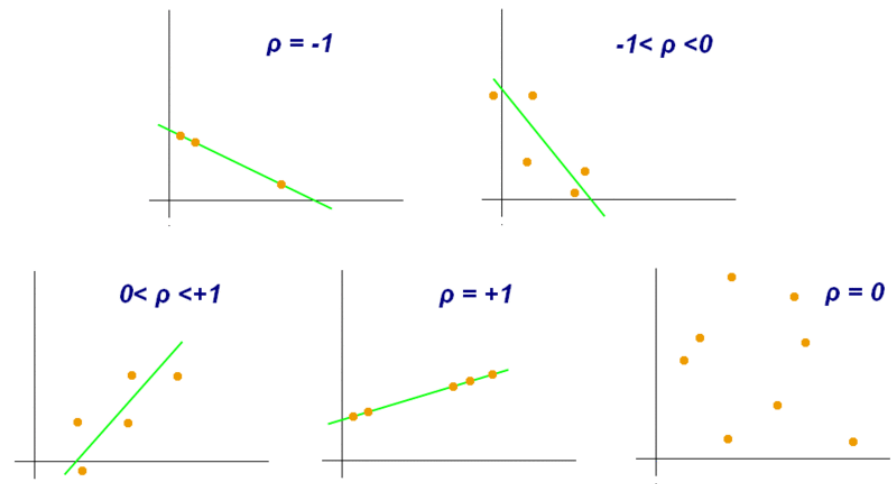


Individual metabolite level (2)

- **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:

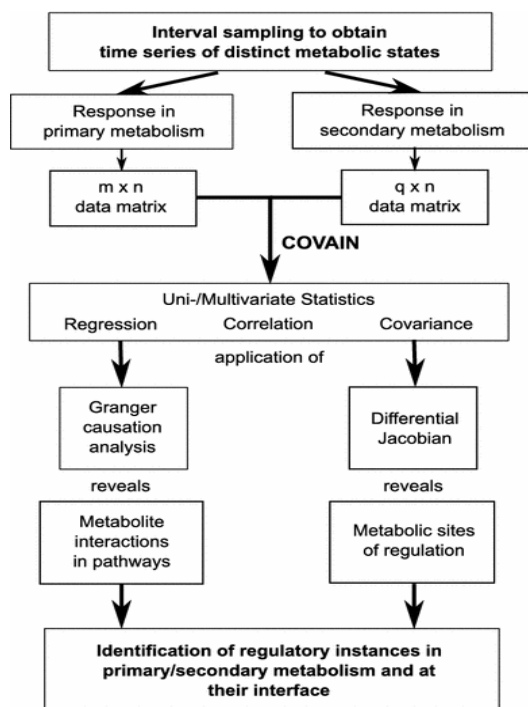
- 1) Correlation coefficients [0-1]
- 2) p-value to indict significance



Individual metabolite level (3)

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

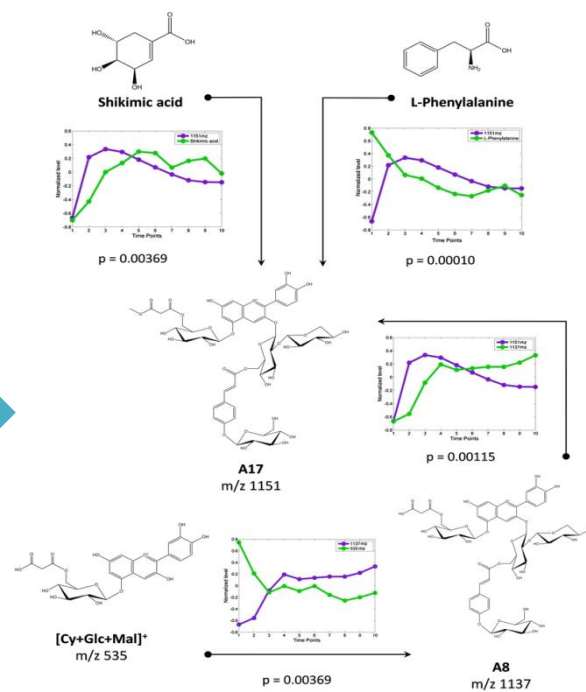
Hannes Doerfler, David Lyon*, Xiaoliang Sun*, ..., Wolfram Weckwerth, Metabolomics, 2012*



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

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

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

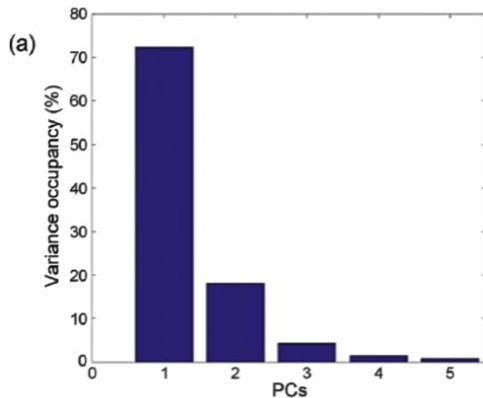


Metabolite group level (1)

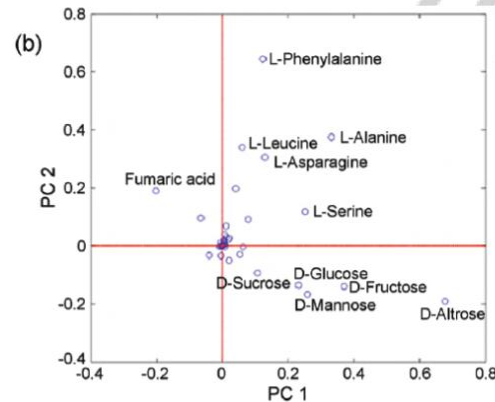
- **Principal component analysis (PCA)** distinguishes phenotypes and finds most influencing metabolites

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

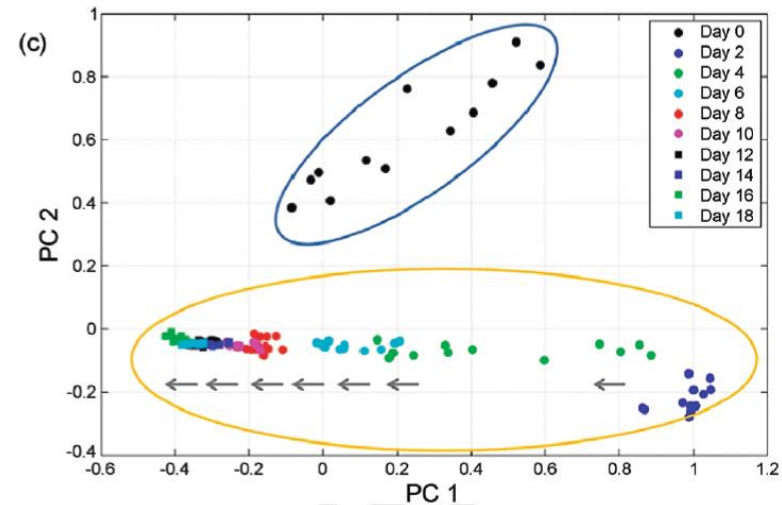
A. Variance occupancy



B. Loading plot



C. Coordinate plot

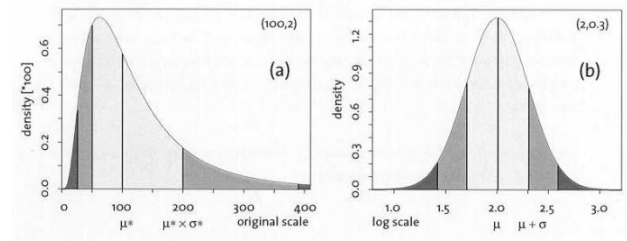
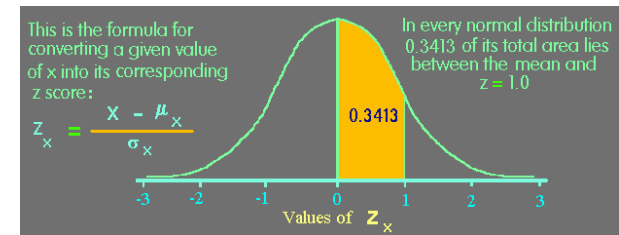


Data Scaling?

- 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

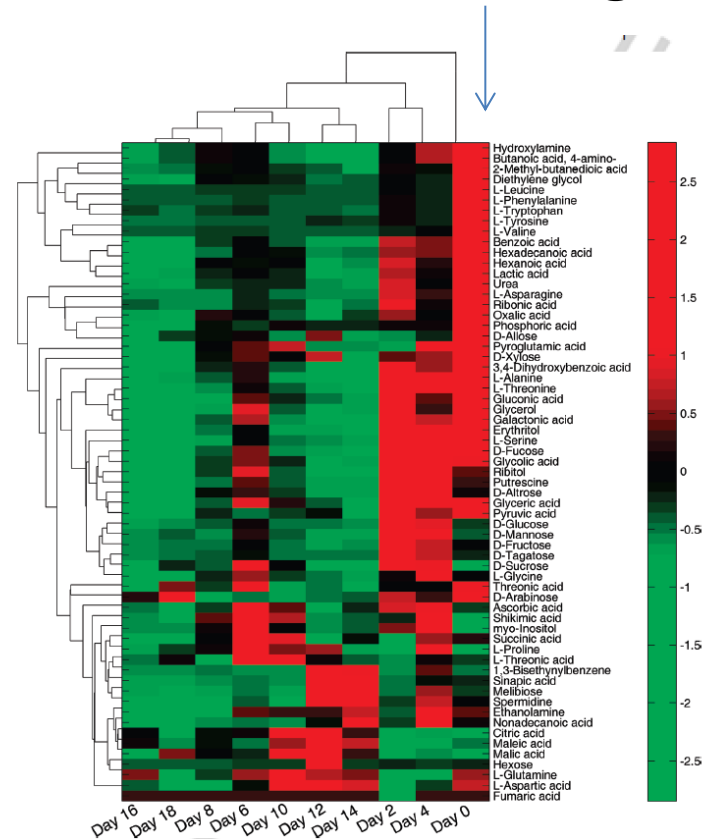
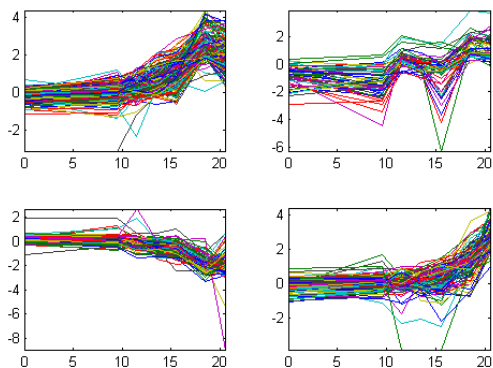
- Scaling methods:
 - Linear: z-score, range [0,1]
 - Nonlinear: log



Metabolite group level (2)

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

K-Means Clustering of Profiles



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-

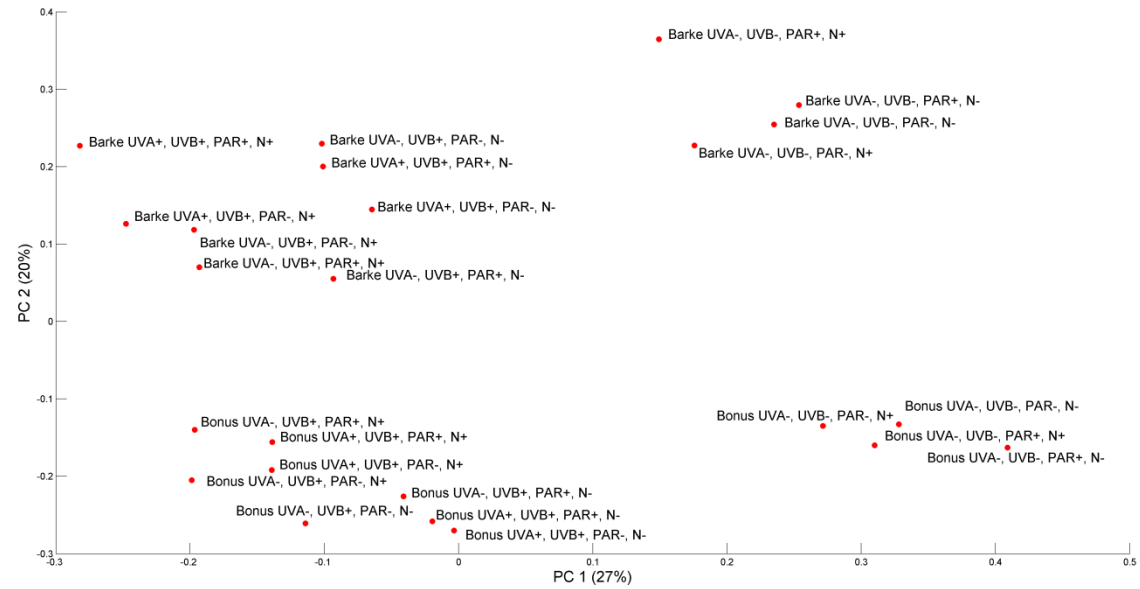
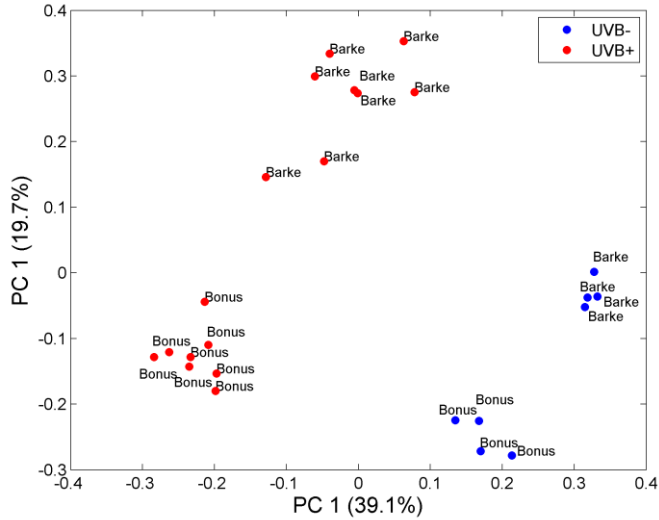
Discretize treatments: 1 for -, 2 for +

Regression equation:

$$\text{m/z feature level} = a_0 + a_1 * \text{Species} + a_2 * \text{Tr.1} + a_3 * \text{Tr.2} + a_4 * \text{Tr.3} + a_5 * \text{Tr.4} + a_6 * \text{Species} * \text{Tr.1} + a_7 * \text{Species} * \text{Tr.2} + a_8 * \text{Species} * \text{Tr.3} + a_9 * \text{Tr.4} + a_{10} * \text{Tr.1} * \text{Tr.2} + a_{11} * \text{Tr.1} * \text{Tr.3} + a_{12} * \text{Tr.1} * \text{Tr.4} + a_{13} * \text{Tr.2} * \text{Tr.3} + a_{14} * \text{Tr.2} * \text{Tr.4} + a_{15} * \text{Tr.3} * \text{Tr.4} + \text{error}$$

Stepwise regression eliminates insignificant associations

Regression results vs. whole data

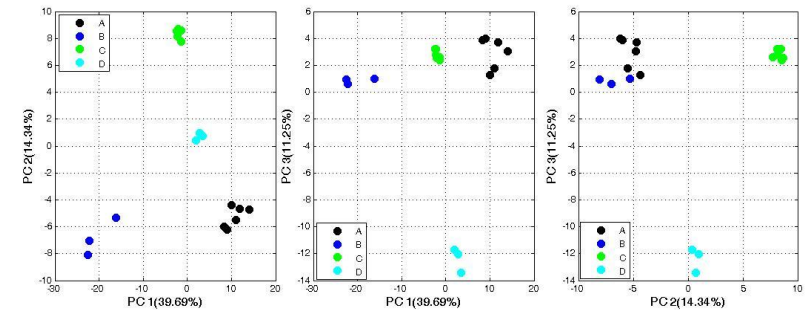


Influences of data preprocessing to statistical analysis results

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	●●	○	●	●	●	○	○

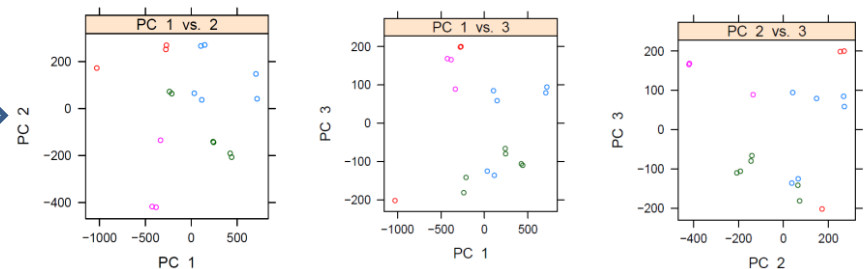
The distribution-based imputation method: using Expectation-Maximization algorithm

COVAIN results



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

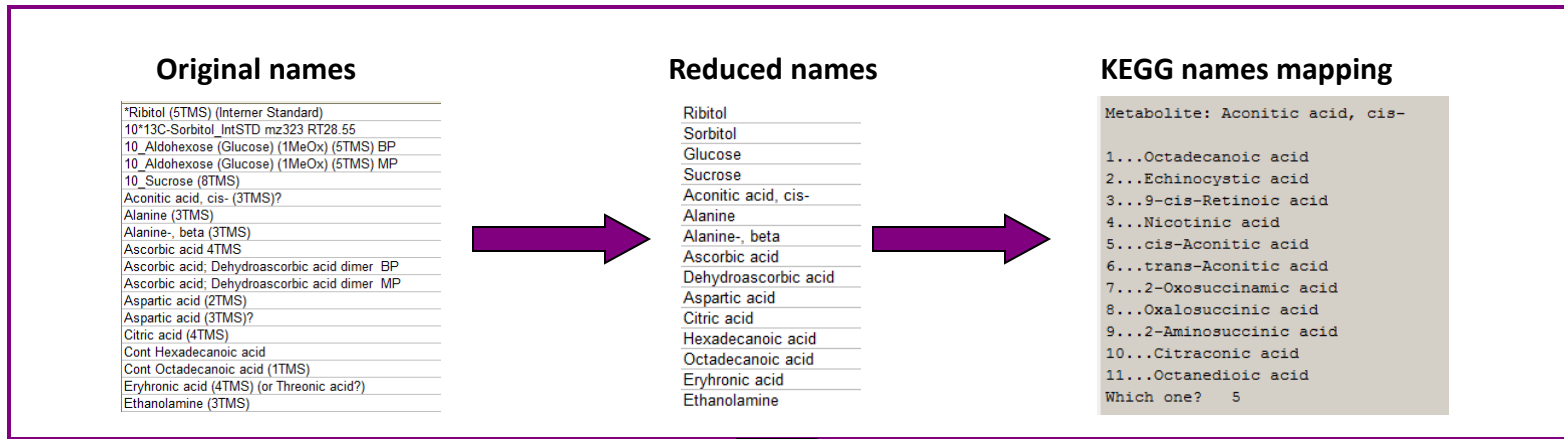
MetaGeneAnalyse results



PCA analysis on a proteomics dataset



Network level (1): database mapping



Original names

*Ribitol (5TMS) (Internal 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)
 Erythronic 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
 Erythronic 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

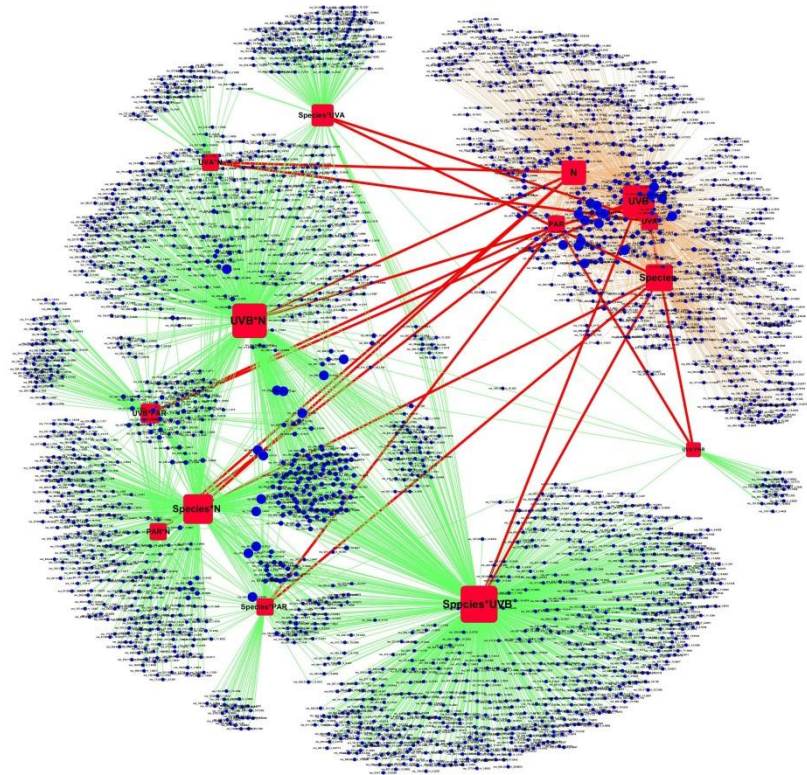
KEGG ID/Formula Mapping

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	5.35646654	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

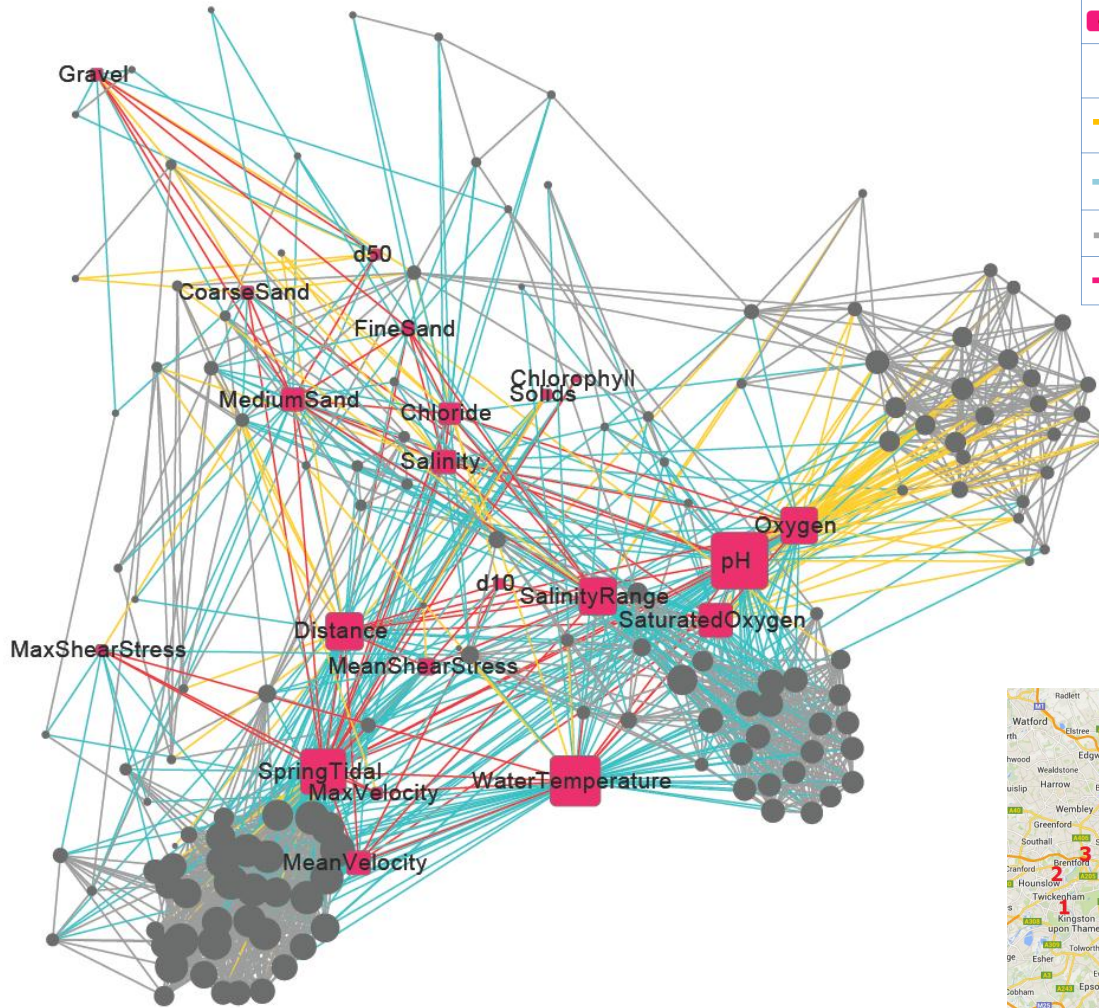
← User's data integration

Network level (2): inferring from data

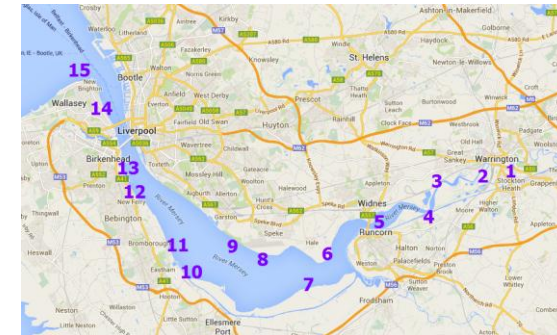
- Based on:
 - Correlation coefficients
 - Granger causality
 - Regression



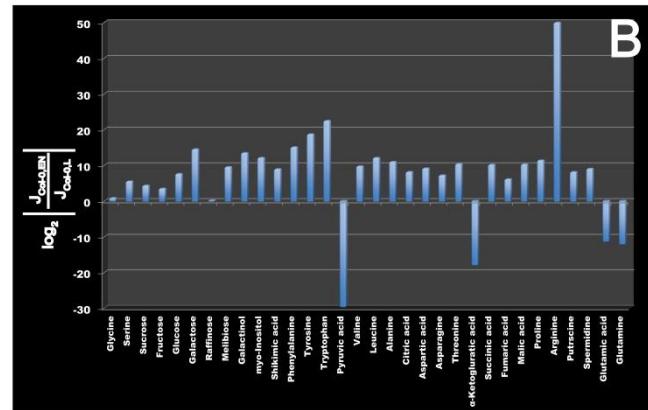
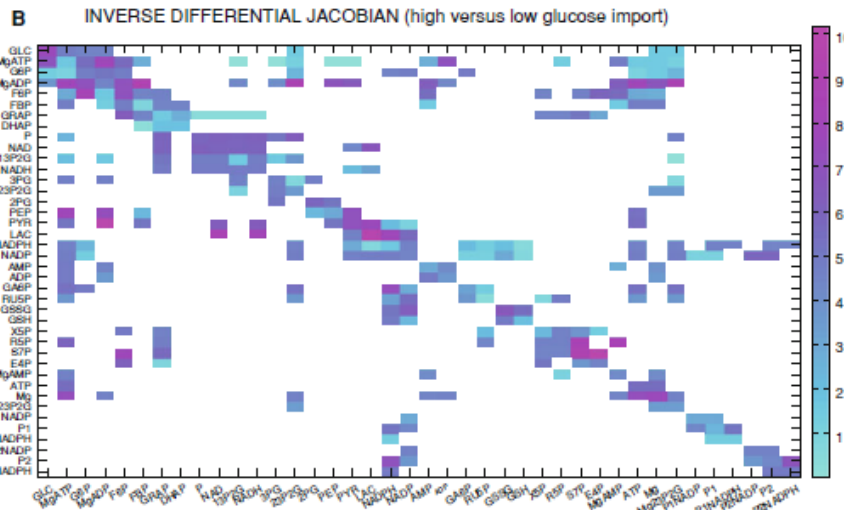
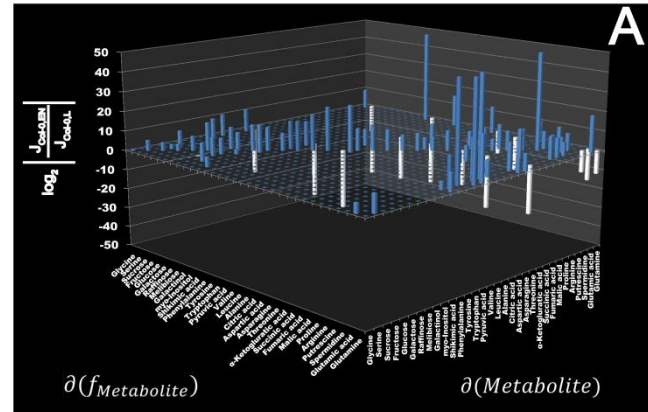
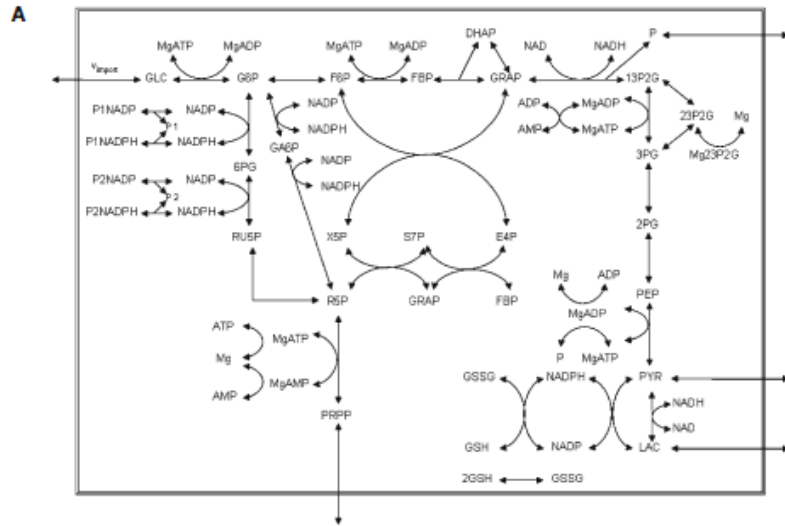
An ecological network



d10	pH	Environmental factors. Size proportional to degree.
•	●	OTUs. Size proportional to degree.
— (yellow)		Associations between OTU and one environmental factor.
— (cyan)		Associations between OTU and multiple environmental factors.
— (grey)		Associations between OTUs.
— (pink)		Associations between environmental factors.

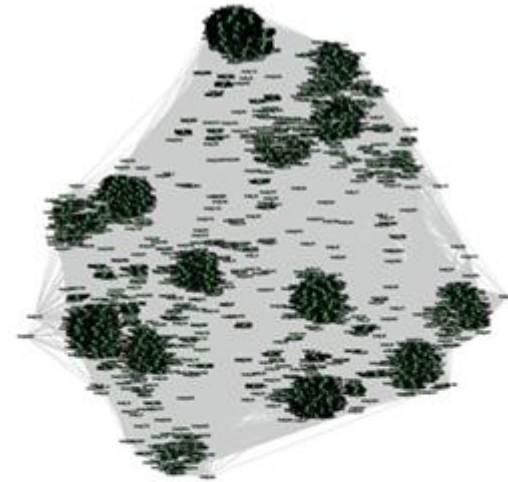
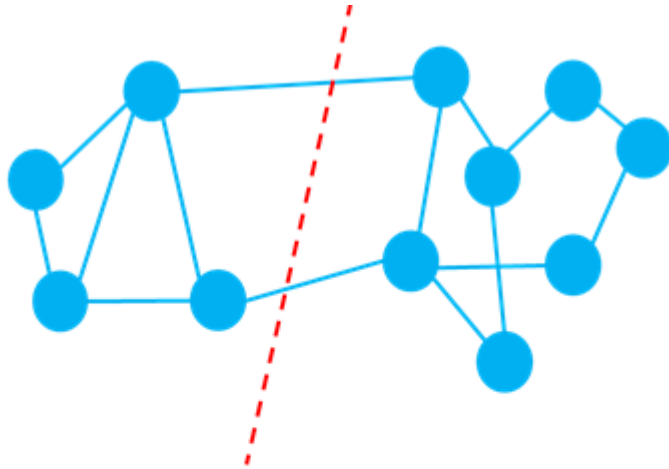


Network level (3): differential Jacobian identifies dynamical changes in metabolic network

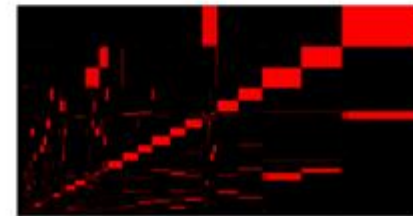


Sun X, Weckwerth W, Metabolomics, 2012
 Nagele T, Mair A, Sun X, et al, PLoS One, 2014

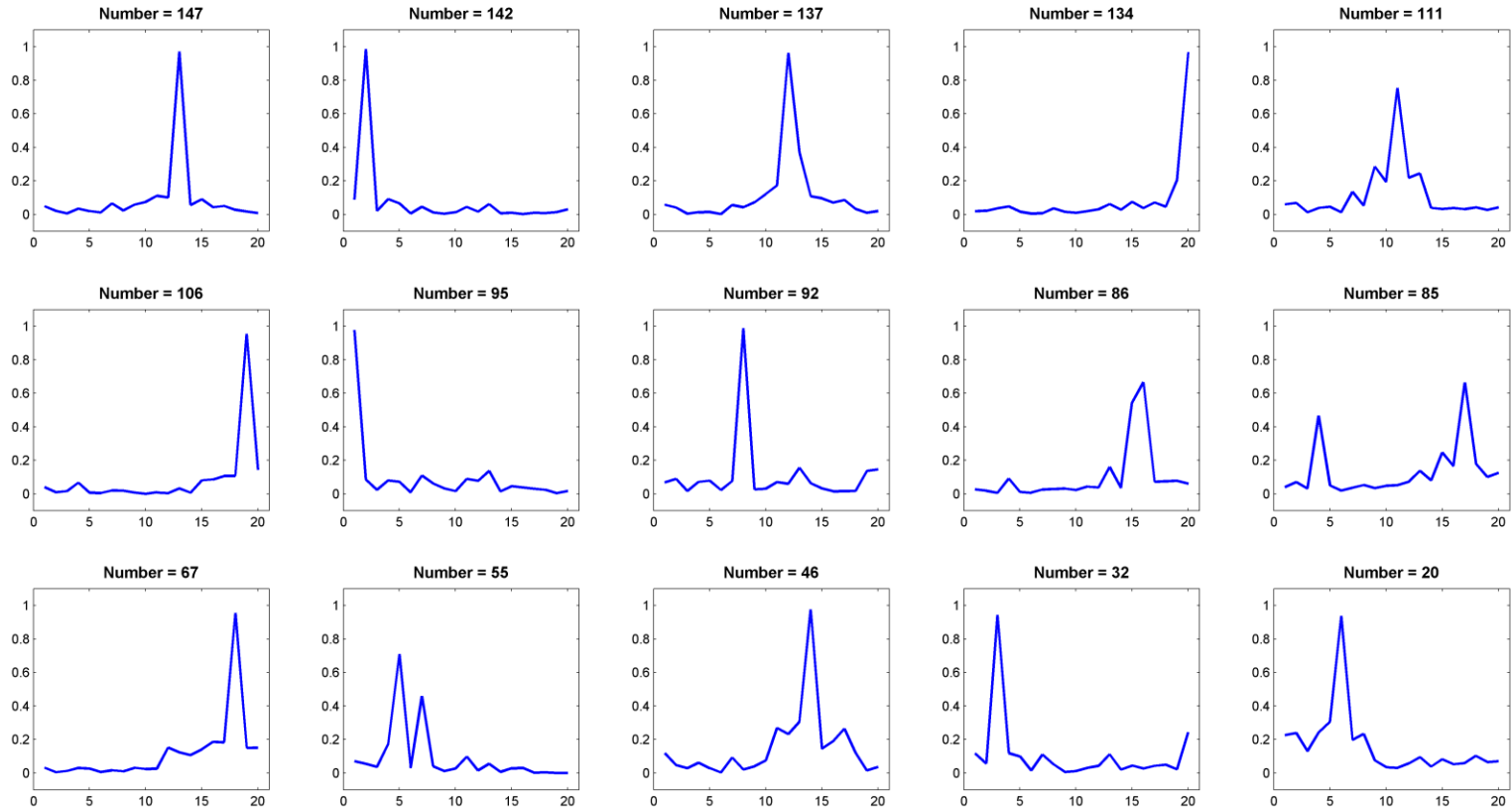
Network level (4): spectral clustering partitions network into separate modules



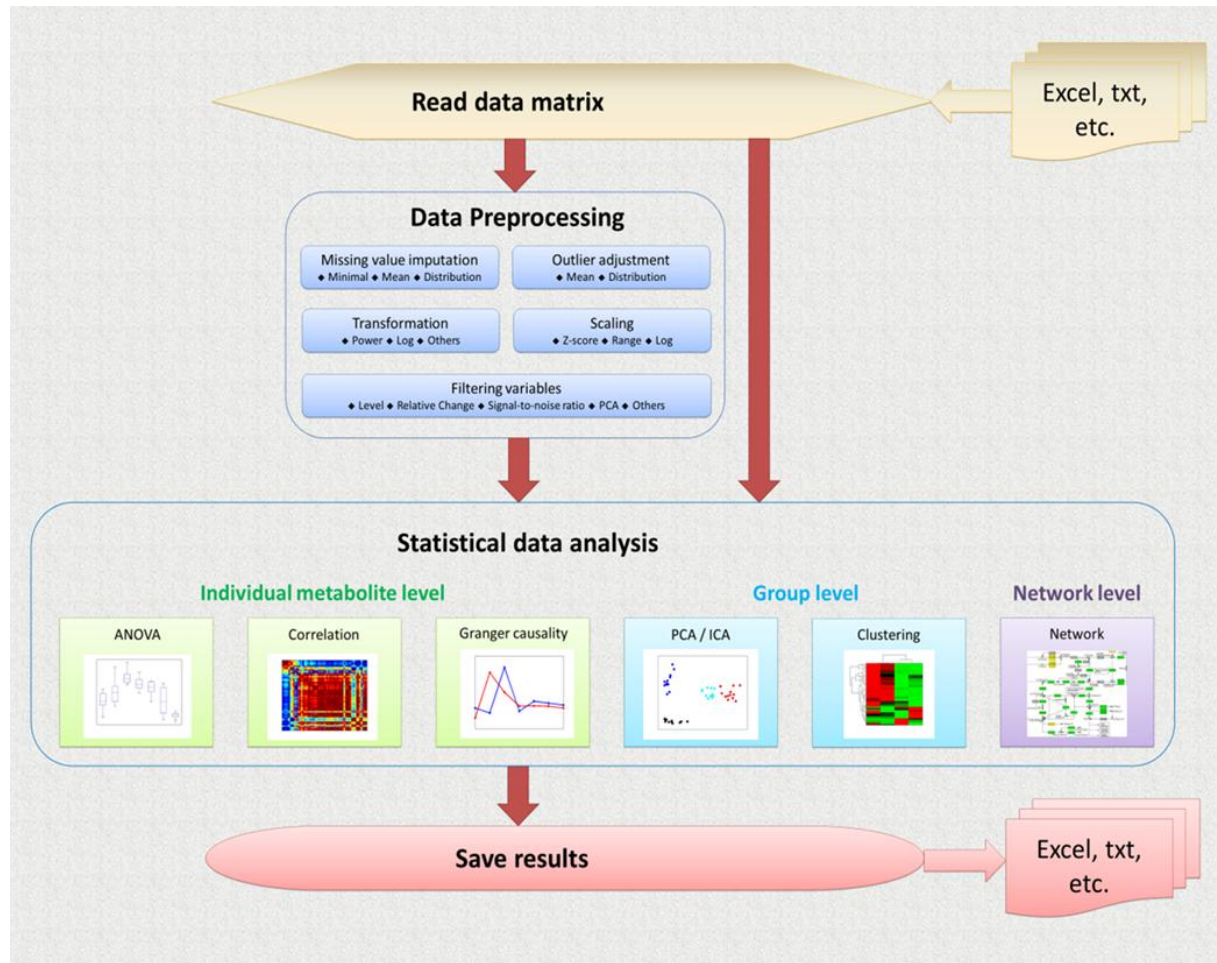
The network is constructed from regression coefficients and is partitioned into a few modules by spectral clustering.



Clustering variables from network



Analyzing metabolomics data by COVAIN



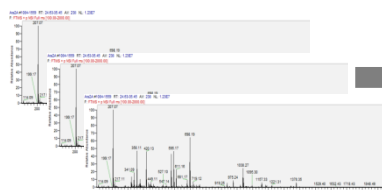
Sun X & Weckwerth W, Metabolomics, 2012
Doerfler H*, Sun X*, et al, PLoS One, 2014

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

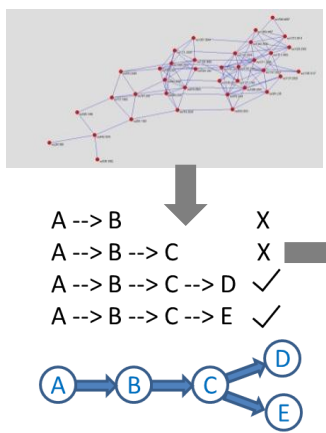
Read m/z data and reaction rules



A	B	C	D	E	F
m/z	Intensity	Relative	Th. Mass	Delta (ppm)	Composition
448.11559	25978.6	0.86	448.11578	-0.41	C11 H22 O14 N5
449.10783	188211.2	6.2	448.11526	0.73	C24 H34 O3 N7
594.18077	9240.6	0.3	449.10784	-0.02	C11 H21 O11
594.35642	8594.1	0.28	594.19028	-0.87	C32 H34 O16 N2
595.14474	71219.1	2.35	594.35644	-0.04	C31 H46 O3 N4
			594.35695	-0.89	C33 H48 O9 N9
			594.35696	-0.9	C42 H54 O14 N2
			595.14461	0.21	C30 H27 O13
			595.14512	-0.64	C16 H29 O19 N5

Reaction step / chemical modification	A	B	C	D	E
	net C change	net H change	net O change	net N change	
methylation / CH2 elongation... (+CH2)	1	2	0	0	
2-methylation / N-methylation... (+CH3)	1	3	0	0	
2-methylthio / N-methylation... (+SCH3)	0	2	0	0	
monooxygenation (+O)	0	0	1	0	
oxidation (+O2)	0	0	2	0	
hydrogenation (+H2)	0	2	0	0	
2-hydroxylation	0	0	4	0	
X-H-X-cysteine system addition	3	5	2	1	
X-H-X-alanine addition	3	5	2	1	
sulfation	0	0	3	0	
combined hydrogenation and protonation	0	1	0	0	
protonation	0	1	0	0	
hydration	0	2	1	0	
z-hydroxylation	0	0	4	2	
oxidoreduction	0	-2	1	0	
heosylation (glycosylation...)	6	10	5	0	
glycosylation (-O) / s-mannosylation	6	10	4	0	

mzGroupAnalyzer + Pathway Viewer



Outputs

Reaction	From	Reaction	To	From	Reaction	To	From	Reaction	To	From	Reaction	To
1	448.11559	449.10783	449.10783	449.10783	449.10783	449.10783	449.10783	449.10783	449.10783	449.10783	449.10783	449.10783
2	449.10783	594.18077	594.18077	594.18077	594.18077	594.18077	594.18077	594.18077	594.18077	594.18077	594.18077	594.18077
3	594.18077	594.35642	594.35642	594.35642	594.35642	594.35642	594.35642	594.35642	594.35642	594.35642	594.35642	594.35642
4	594.35642	595.14474	595.14474	595.14474	595.14474	595.14474	595.14474	595.14474	595.14474	595.14474	595.14474	595.14474

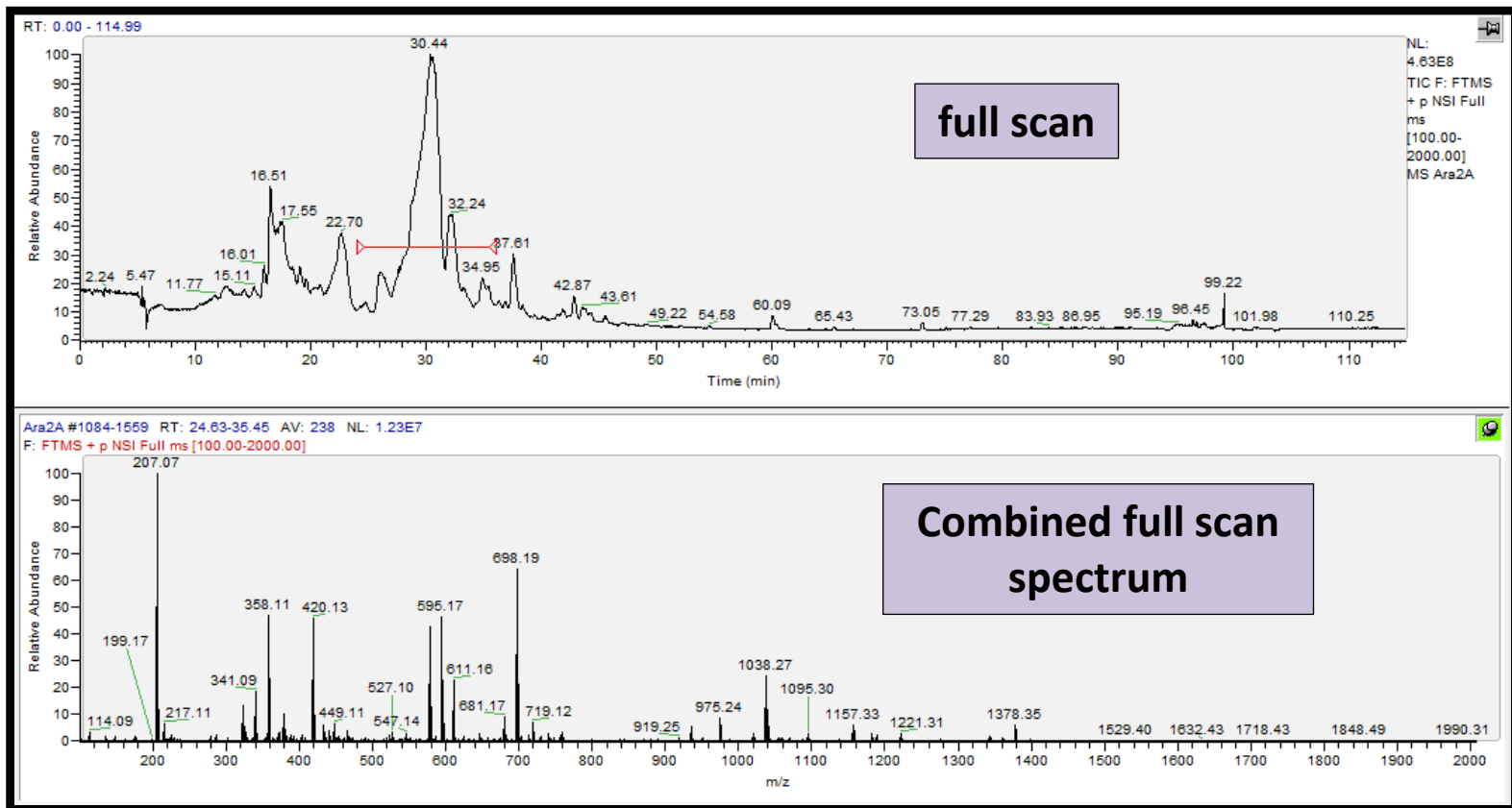
Case	Name	DFC	DFH	DFO	DFN	DFR
43	hydroxylation (hydroxylation...)	6	10	5	0	
88	methylation	3	2	3	0	
17	z-hydroxylation	0	4	0	0	
18	monooxygenation (+O)	0	0	1	0	
24	oxidoreduction	9	6	2	0	
30	hydroxylation (+H2)	2	4	2	0	
32	carbonation	11	10	4	0	
33	hydroxylation (+O)	0	2	0	0	
34	hydroxylation (+O)	0	2	0	0	
35	hydroxylation (+O)	0	2	0	0	
36	hydroxylation (+O)	0	2	0	0	
37	hydroxylation (+O)	0	2	0	0	
38	hydroxylation (+O)	0	2	0	0	
39	hydroxylation (+O)	0	2	0	0	
40	hydroxylation (+O)	0	2	0	0	
41	hydroxylation (+O)	0	2	0	0	
42	hydroxylation (+O)	0	2	0	0	
43	hydroxylation (+O)	0	2	0	0	
44	hydroxylation (+O)	0	2	0	0	
45	hydroxylation (+O)	0	2	0	0	
46	hydroxylation (+O)	0	2	0	0	
47	hydroxylation (+O)	0	2	0	0	
48	hydroxylation (+O)	0	2	0	0	
49	hydroxylation (+O)	0	2	0	0	
50	hydroxylation (+O)	0	2	0	0	
51	hydroxylation (+O)	0	2	0	0	
52	hydroxylation (+O)	0	2	0	0	
53	hydroxylation (+O)	0	2	0	0	
54	hydroxylation (+O)	0	2	0	0	
55	hydroxylation (+O)	0	2	0	0	
56	hydroxylation (+O)	0	2	0	0	
57	hydroxylation (+O)	0	2	0	0	
58	hydroxylation (+O)	0	2	0	0	
59	hydroxylation (+O)	0	2	0	0	
60	hydroxylation (+O)	0	2	0	0	
61	hydroxylation (+O)	0	2	0	0	
62	hydroxylation (+O)	0	2	0	0	
63	hydroxylation (+O)	0	2	0	0	
64	hydroxylation (+O)	0	2	0	0	
65	hydroxylation (+O)	0	2	0	0	
66	hydroxylation (+O)	0	2	0	0	
67	hydroxylation (+O)	0	2	0	0	
68	hydroxylation (+O)	0	2	0	0	
69	hydroxylation (+O)	0	2	0	0	
70	hydroxylation (+O)	0	2	0	0	
71	hydroxylation (+O)	0	2	0	0	
72	hydroxylation (+O)	0	2	0	0	
73	hydroxylation (+O)	0	2	0	0	
74	hydroxylation (+O)	0	2	0	0	
75	hydroxylation (+O)	0	2	0	0	
76	hydroxylation (+O)	0	2	0	0	
77	hydroxylation (+O)	0	2	0	0	
78	hydroxylation (+O)	0	2	0	0	
79	hydroxylation (+O)	0	2	0	0	
80	hydroxylation (+O)	0	2	0	0	
81	hydroxylation (+O)	0	2	0	0	
82	hydroxylation (+O)	0	2	0	0	
83	hydroxylation (+O)	0	2	0	0	
84	hydroxylation (+O)	0	2	0	0	
85	hydroxylation (+O)	0	2	0	0	
86	hydroxylation (+O)	0	2	0	0	
87	hydroxylation (+O)	0	2	0	0	
88	hydroxylation (+O)	0	2	0	0	
89	hydroxylation (+O)	0	2	0	0	
90	hydroxylation (+O)	0	2	0	0	
91	hydroxylation (+O)	0	2	0	0	
92	hydroxylation (+O)	0	2	0	0	
93	hydroxylation (+O)	0	2	0	0	
94	hydroxylation (+O)	0	2	0	0	
95	hydroxylation (+O)	0	2	0	0	
96	hydroxylation (+O)	0	2	0	0	
97	hydroxylation (+O)	0	2	0	0	
98	hydroxylation (+O)	0	2	0	0	
99	hydroxylation (+O)	0	2	0	0	
100	hydroxylation (+O)	0	2	0	0	

```

data: [489x5 double]
dataAll: [489x5 cell]
finalCorrectPath: [3329x1 cell]
finalTimePath: [3329x1 cell]
mmNet: [489x489 logical]
mmNetS: [489x489 double]
mmNetPattern: [489x489 double]
mList: [489x1 double]
mName: [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.:

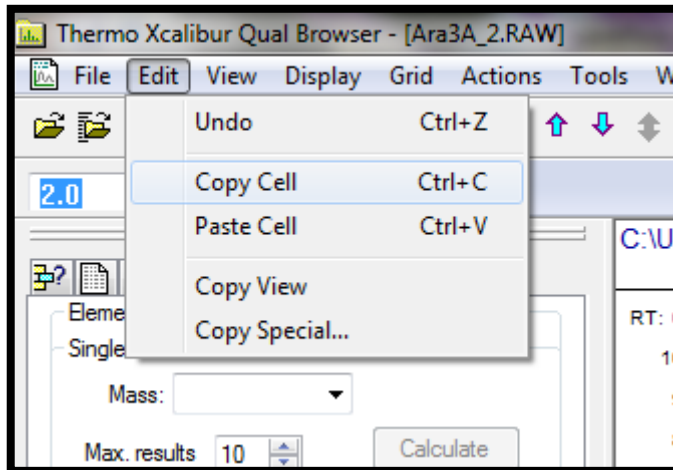
- elements C,H,O,N,S
- Intensity 0.5-100 %
- 5 ppm, 1-2 or lower for high res instruments
- 5 formulae
- ...

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)“

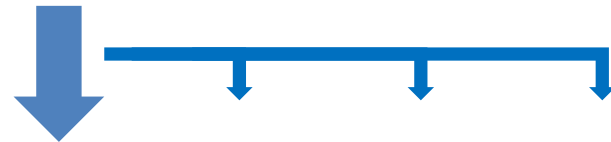


The screenshot shows the 'Export' menu in the Thermo Xcalibur Qual Browser. The menu items are: View, Subtract Spectra, Export, Elemental Comp. (checked), Ranges..., and Display Options... The 'Export' sub-menu is open, showing 'Clipboard (Exact Mass)' and 'Clipboard (Nominal Mass)'. The background shows a table with columns for 'Mass', 'Delta (ppm)', and 'Composition'.

Mass	Delta (ppm)	Composition
9.02332	0.10	C ₈ H ₅ O ₃
9.03858	-15.15	C ₁₂ H ₅
9.00806	15.36	C ₄ H ₅ O ₆
9.04445	-21.02	C ₅ H ₉ O ₅
9.00219	21.23	C ₁₁ H ₁₁ O
7.06519	0.03	C ₁₁ H ₁₁ O ₄
7.08044	-15.23	C ₁₅ H ₁₁ O
7.04993	15.28	C ₇ H ₁₁ O ₇
7.08631	-21.10	C ₈ H ₁₅ O ₆

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

Data information

Choose data

full data preview
 mean value preview
 transpose full preview
 transpose mean preview

COVAIN

guided user interface

Load data ...

Combine data ...

My notes ...

Name the results

Save

Data Analysis

Multivariate Statistics

Data selection

Input index, such as 1,2,3 All

PCA

PC number (max 5)

ICA

IC number (max 5)

Correlation

Cluster

Time Series

Time points selection

Input index, such as 4,2,3,1 Default

Correlation

Clustering

Hierarchical K-Means

Granger causation analysis

Permutation Entropy

Network Analysis

Network Inference

Network Property

Inverse Jacobian

KEGG Pathway

Options

Help

1. load data

2. select data, rule file and output location

Index	Value	Value	Value	Value	Value
21	594.35642	8594.1	0.28	334.35644	-0.04 C17 H46 O1 N4
22	594.35695			594.35695	-0.89 C23 H48 O9 N9
23	594.36094			594.36094	-0.9 C24 H48 O10 N2
24	595.14474	71215.1	2.35	595.14442	0.31 C10 H17 O13
25	595.14461			595.14461	0.22 C19 H21 O8 N7
26	595.14512			595.14512	-0.64 C16 H23 O13 N5
27	595.18563	84340.1	23.73	595.18574	-0.18 C20 H23 O15 N7
28	595.18575			595.18575	-0.19 C27 H41 O13
29	595.18523			595.18523	0.67 C40 H23 O4 N2
30	595.3688	12341.7	0.41	595.3688	0 C28 H47 O17 N7
31	595.3688			595.3688	-0.01 C29 H53 O12
32	727.18893	20668.8	0.66	727.18869	0.85 C47 H23 D N9
33	727.18867			727.18867	0.08 C34 H29 O13 N7
34	727.18778			727.18778	-0.62 C31 H17 O23 N3
35	727.18836			727.18836	0.78 C48 H27 O4 N2
37	727.18836			727.18836	0.79 C47 H23 D N9
38	727.18756			727.18756	0.28 C40 H48 O8 N2
39	727.35807			727.35807	-0.42 C38 H53 O14 N10
40	727.35806			727.35806	-0.43 C27 H17 O23 N3
41	727.82506			727.82506	0.97 C34 H47 O2
42	728.19048	7156.8	0.24	728.19047	-0.26 C19 H19 O14 N9
43	728.19046			728.19046	-0.27 C10 H19 O13 N2
44	728.19017			728.19017	0.43 C43 H28 O4 N2
45	728.18984			728.18984	0.67 C14 H18 O14 N10

data file

+

Index	Value	Value	Value	Value	Value
1	Reaction step / Chemical modification			met mass ch net C change net H change net O change	
2	methylation / C10 alkylation... (+CH2)			14.0156501	1
3	methylation / N-methylation... (+CH3)			15.0208751	1
4	demethylation / C10 alkylation... (-2CH2)			28.0113001	2
5	demethylation / N-methylation... (-2CH3)			30.0485022	2
6	monooxygenation (+2HCHO)			15.0495446	0
7	oxidation (+O)			11.9898293	0
8	hydrogenation (+H2)			1.1216407	0
9	dehydrogenation (-H2)			4.6113013	0
10	combined hydrogenation and protonation			3.0214711	0
11	protonation			1.0079204	0
12	hydration			18.0205647	0
13	dehydration			36.0212394	0
14	oxidoreduction			11.9792848	0
15	hydroxylation (hydroxylation...)			161.053024	6
16	glyoxylation (C1 / aldehyde)			146.0277909	6
17	glyoxalation (2O)			130.062394	6
18	peroxidation (hydroxylation, ribosylation...)			133.042039	5
19	malonylation			86.000394	3
20	coumaroylation			148.049376	9
21	unsaturation			266.057909	11
22	carboxylation			43.9898293	1
23	hydroxylation			156.0208751	7
24	acetylation			43.0205647	2
25	A) methylation B) amino acid formation from aspartic C) C14 "insertion" after			30.0205647	1

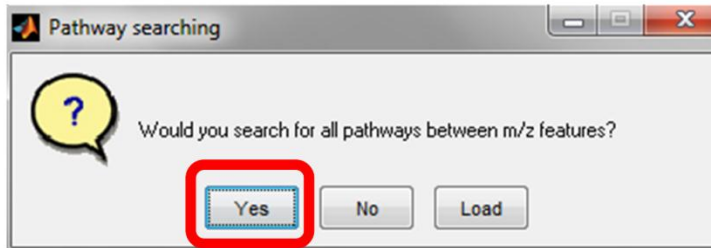
rule file

+

output directory

→ transformation files will be created

3. start *Pathway Viewer*



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



mzGroupAnalyzer Pathway Viewer

Select & List

Composition: C, H, O, TP 1

m/z value: From, to

Chemical Transformation Rules

- Select / Reject all
- methylation / CH2 elongation... (+CH2)
- methylation / N-methylation... (+CH3)
- 2xmethylation / CH2 elongation... (+ 2x...)
- 2xmethylation / N-methylation... (+2xC...)
- monooxygenation (+O)
- oxidation (+O2)
- hydrogenation (+H2)
- 2hydrogenation
- combined hydrogenation and protonati...
- protonation
- hydration
- 2hydration
- oxidoreduction
- hexosylation (glycosylation,...)
- glycosylation (-O) / rhamnosylation
- glycosylation (-2O)
- pentosylation (xylosylation, ribosylation...)
- malonylation
- coumaroylation
- sinapoylation
- carboxylation
- benzoylation
- acetylation
- A) methoxylation B) -amino acid format...
- dihydroxylation of double bond
- reduction (-O+H)
- reduction (-2O+2H)
- glucuronidation
- carboxylation+hydrogenation
- C=O addition
- dimethyl allyl addition
- hydration + protonation
- amination
- amidation
- H2O addition under proton loss
- formic acid methyl ester formation
- oxidative deamination of amin
- formylation
- vicinal diol insertion
- X-H->X-cysteine cystein addition
- X-H-> X-alanine addition
- sulfation
- C=O->C=N dihydroxy phenylalanin add...
- C=O->C=N phenylalanin addition
- 3-methoxy-p-coumaric acid addition
- betalamic acid addition
- glyceraldehyde addition
- amidino group addition
- reaction with net C loss (e.g. phenylpyr...
- 3-amino-4,7-dihydroxy-8-methylcouma...
- glutamination
- caffeic acid addition
- phosphorylation

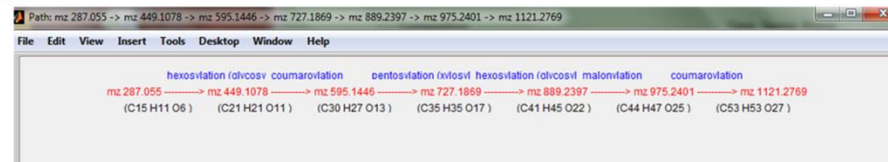
Time Series Points

Show Export

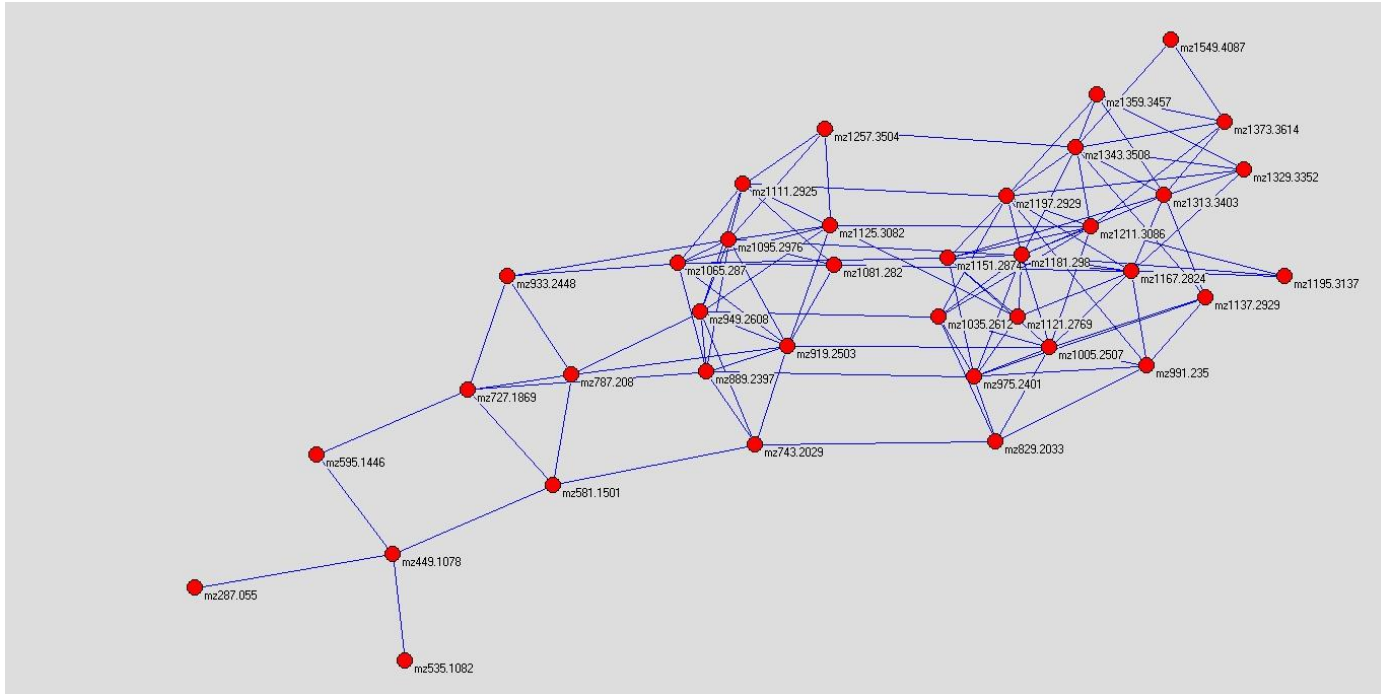
Totally 163 paths are found!

	From (m/z)	From (CHO)	To (m/z)	To (CHO)	Path
1	mz 287.055	C15 H11 O6	mz 535.1082	C24 H23 O...	mz 287.055 -> mz 449.1078 -> mz 535.1082
2	mz 287.055	C15 H11 O6	mz 889.2397	C41 H45 O...	mz 287.055 -> mz 449.1078 -> mz 595.1446 -> mz 889.2397
3	mz 287.055	C15 H11 O6	mz 1005.2...	C45 H49 O...	mz 287.055 -> mz 449.1078 -> mz 581.1501 -> mz 1005.2...
4	mz 287.055	C15 H11 O6	mz 1035.2...	C46 H51 O...	mz 287.055 -> mz 449.1078 -> mz 581.1501 -> mz 1035.2...
5	mz 287.055	C15 H11 O6	mz 1065.287	C51 H53 O...	mz 287.055 -> mz 449.1078 -> mz 595.1446 -> mz 1065.287
6	mz 287.055	C15 H11 O6	mz 1081.282	C51 H53 O...	mz 287.055 -> mz 449.1078 -> mz 581.1501 -> mz 1081.282
7	mz 287.055	C15 H11 O6	mz 1111.2...	C52 H55 O...	mz 287.055 -> mz 449.1078 -> mz 581.1501 -> mz 1111.2...
8	mz 287.055	C15 H11 O6	mz 1121.2...	C53 H53 O...	mz 287.055 -> mz 449.1078 -> mz 581.1501 -> mz 1121.2...
9	mz 287.055	C15 H11 O6	mz 1151.2...	C54 H55 O...	mz 287.055 -> mz 449.1078 -> mz 581.1501 -> mz 1151.2...
10	mz 287.055	C15 H11 O6	mz 1167.2...	C54 H55 O...	mz 287.055 -> mz 449.1078 -> mz 581.1501 -> mz 1167.2...
11	mz 287.055	C15 H11 O6	mz 1195.3...	C56 H59 O...	mz 287.055 -> mz 449.1078 -> mz 581.1501 -> mz 1195.3...
12	mz 287.055	C15 H11 O6	mz 1257.3...	C58 H65 O...	mz 287.055 -> mz 449.1078 -> mz 581.1501 -> mz 1257.3...
13	mz 287.055	C15 H11 O6	mz 1313.3...	C60 H65 O...	mz 287.055 -> mz 449.1078 -> mz 581.1501 -> mz 1313.3...
14	mz 287.055	C15 H11 O6	mz 1329.3...	C60 H65 O...	mz 287.055 -> mz 449.1078 -> mz 581.1501 -> mz 1329.3...
15	mz 287.055	C15 H11 O6	mz 1359.3...	C61 H67 O...	mz 287.055 -> mz 449.1078 -> mz 581.1501 -> mz 1359.3...
16	mz 287.055	C15 H11 O6	mz 1373.3...	C62 H69 O...	mz 287.055 -> mz 449.1078 -> mz 581.1501 -> mz 1373.3...
17	mz 287.055	C15 H11 O6	mz 1549.4...	C72 H77 O...	mz 287.055 -> mz 449.1078 -> mz 581.1501 -> mz 1549.4...
18	mz 581.1501	C26 H29 O...	mz 889.2397	C41 H45 O...	mz 581.1501 -> mz 727.1869 -> mz 889.2397
19	mz 581.1501	C26 H29 O...	mz 1065.287	C51 H53 O...	mz 581.1501 -> mz 727.1869 -> mz 933.2448 -> mz 1065.287
20	mz 595.1446	C30 H27 O...	mz 949.2608	C43 H49 O...	mz 595.1446 -> mz 727.1869 -> mz 787.208 -> mz 949.2608
21	mz 595.1446	C30 H27 O...	mz 991.2397	C44 H47 O...	mz 595.1446 -> mz 727.1869 -> mz 889.2397 -> mz 991.2397
22	mz 595.1446	C30 H27 O...	mz 1005.2...	C45 H49 O...	mz 595.1446 -> mz 727.1869 -> mz 889.2397 -> mz 1005.2...







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.