



evropský
sociální
fond v ČR



EVROPSKÁ UNIE



MINISTERSTVO ŠKOLSTVÍ,
MLÁDEŽE A TĚLOVÝCHOVY



INVESTICE DO ROZVOJE VZDĚLÁVÁNÍ

Principles and Practices on analyzing metabolomics data

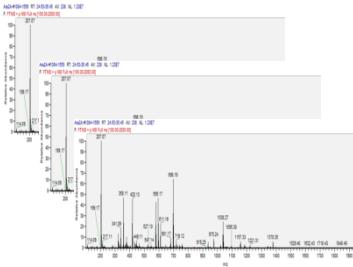
Dr. Xiaoliang Sun

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



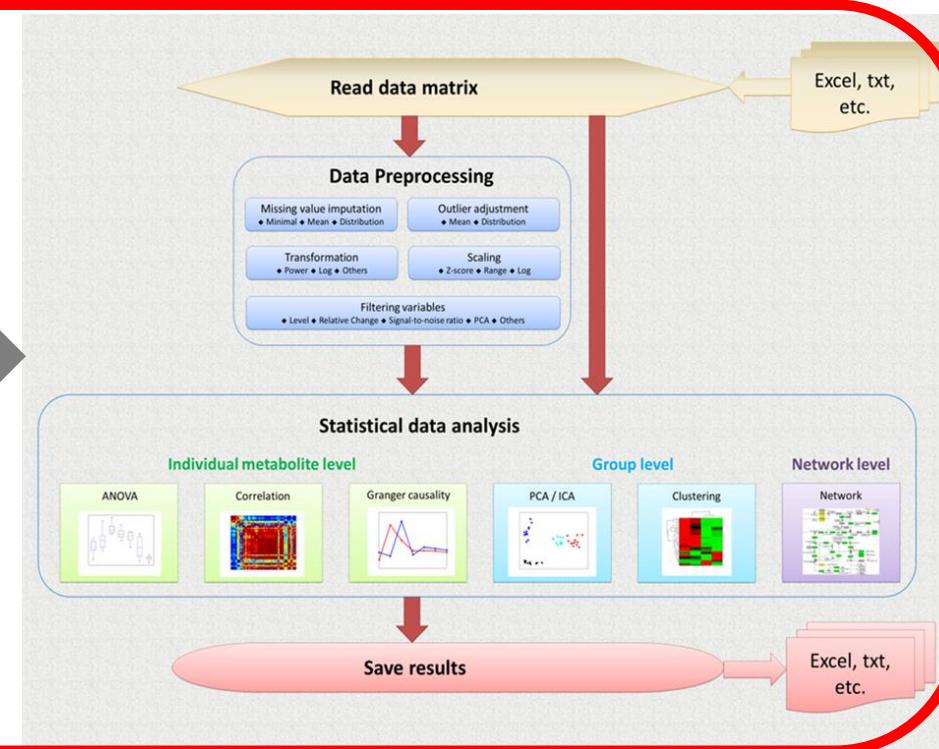
A	B	C	D	E	F
m/z	Intensity	Relative	Theo. Mass	Delta (ppm)	Composition
448.11559	25978.6	0.86	448.11578	-0.41	C11 H22 O14 N5
			448.11527	0.72	C25 H20 O8
			448.11526	0.73	C24 H14 O3 N7
449.10783	188211.2	6.2	449.10783	-0.01	C20 H15 O6 N7
			449.10784	-0.02	C21 H21 O11
594.18977	9240.6	0.3	594.18977	-0.01	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	5.31865579	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.)

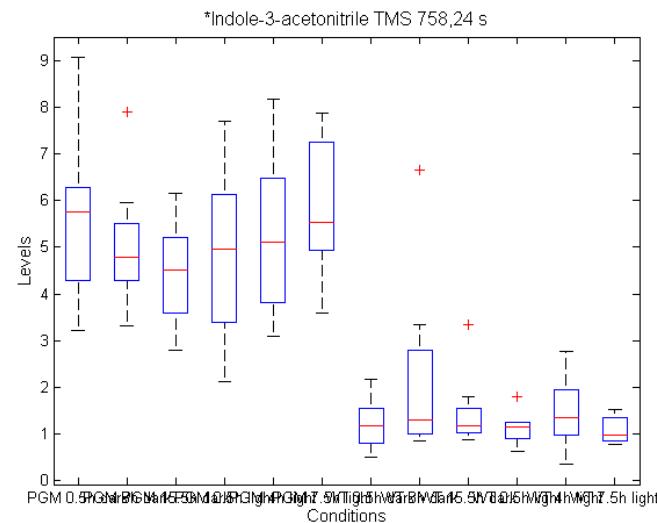
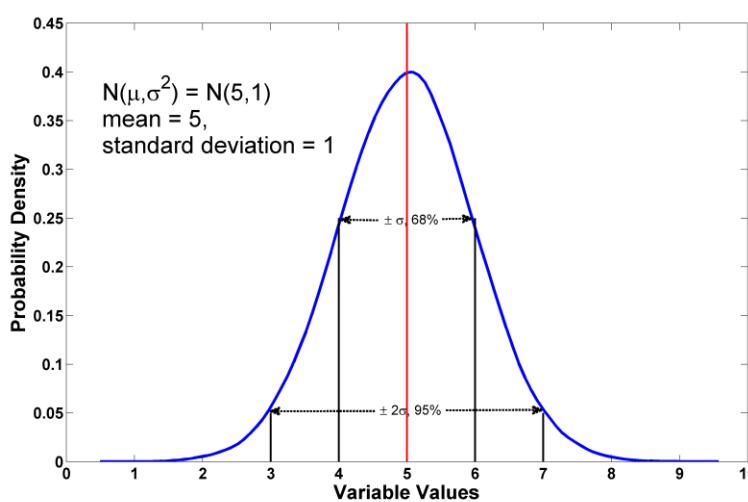
Principles and methods to analyze metabolomics data

- 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

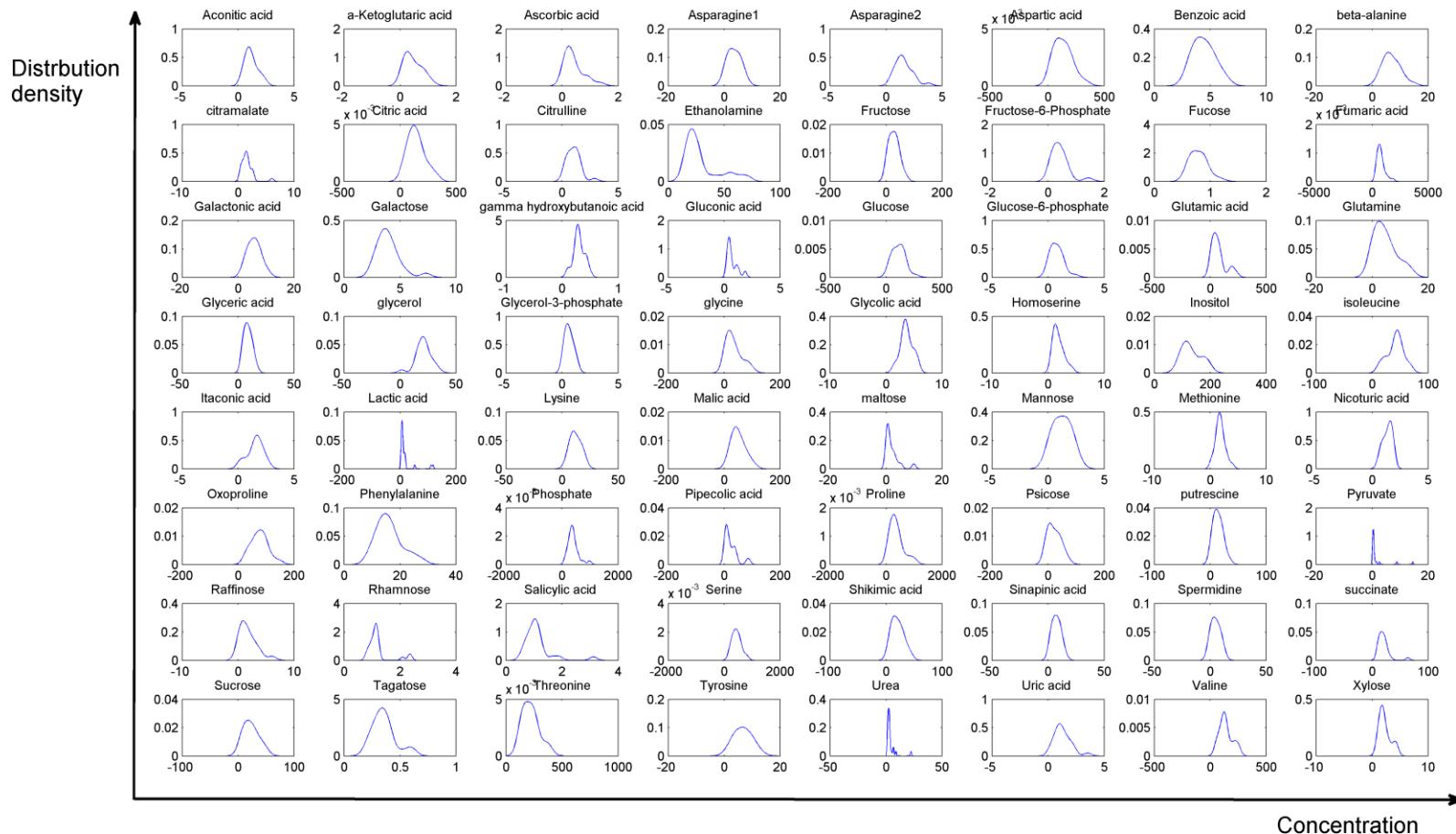
Individual metabolite level (1)

- 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



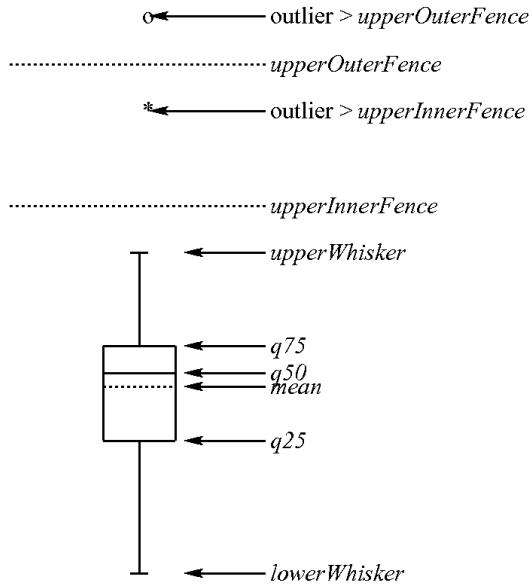
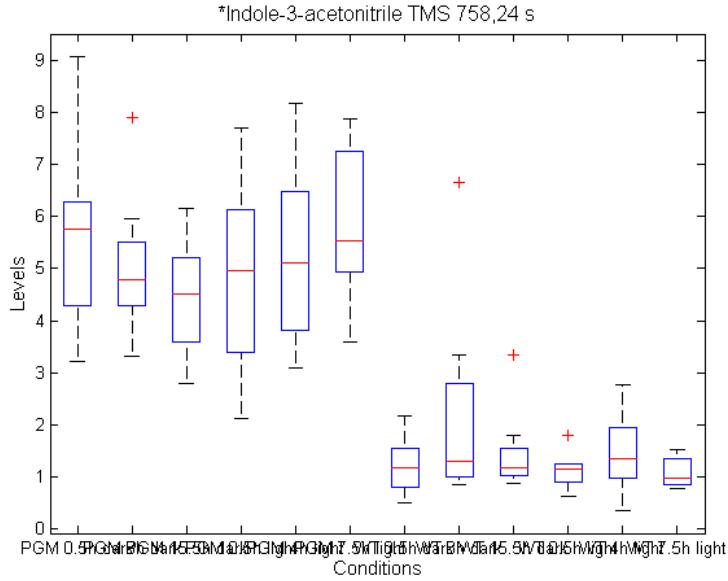
A metabolomics dataset



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

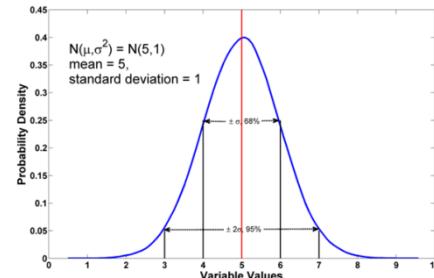
Outliers



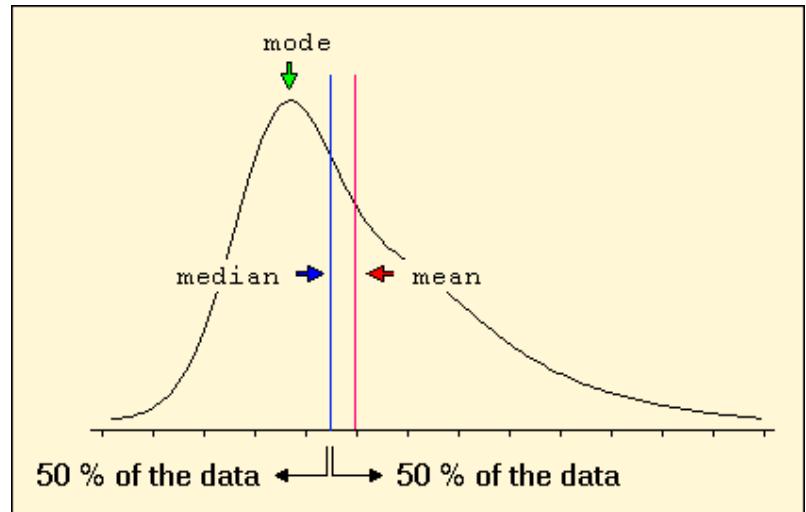
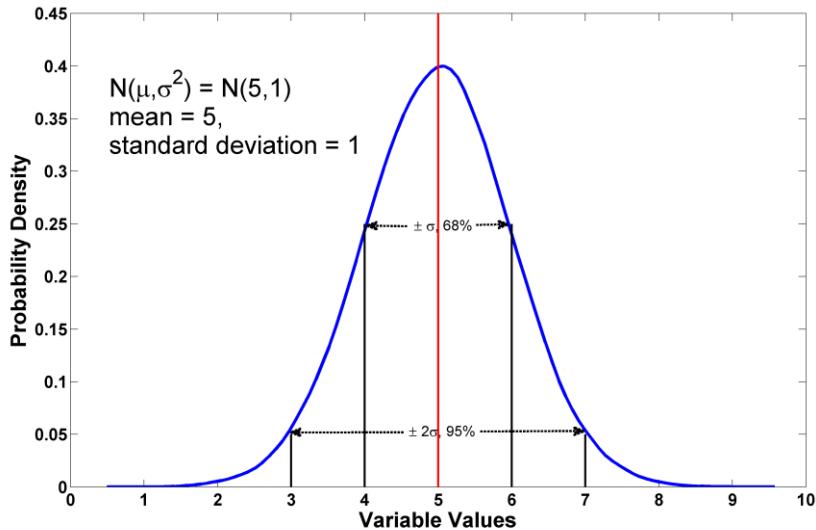
Must consider: are there biological explanations?

Did you know Marie Curie's story?

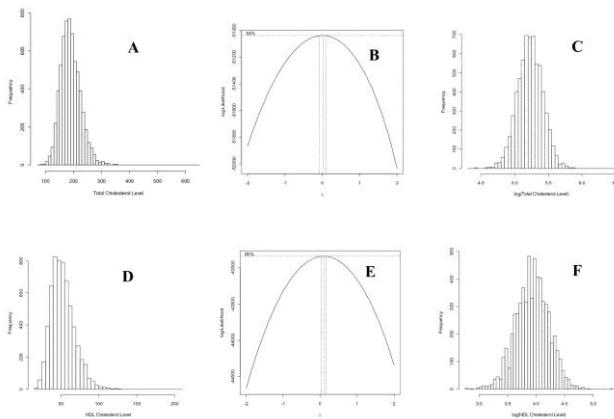
Adjust outliers from the proposed distribution



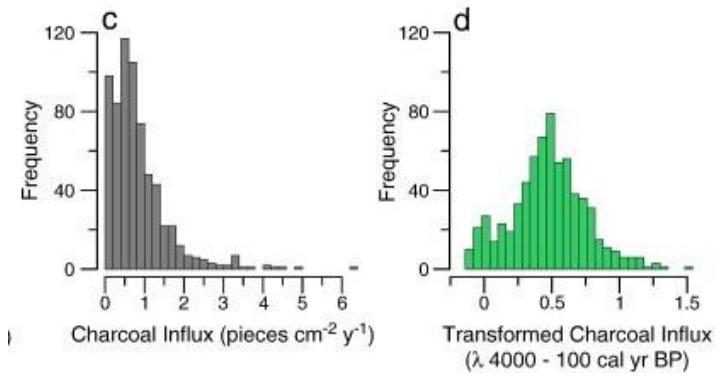
Not normally distributed data



Log transformation



Box-cox transformation

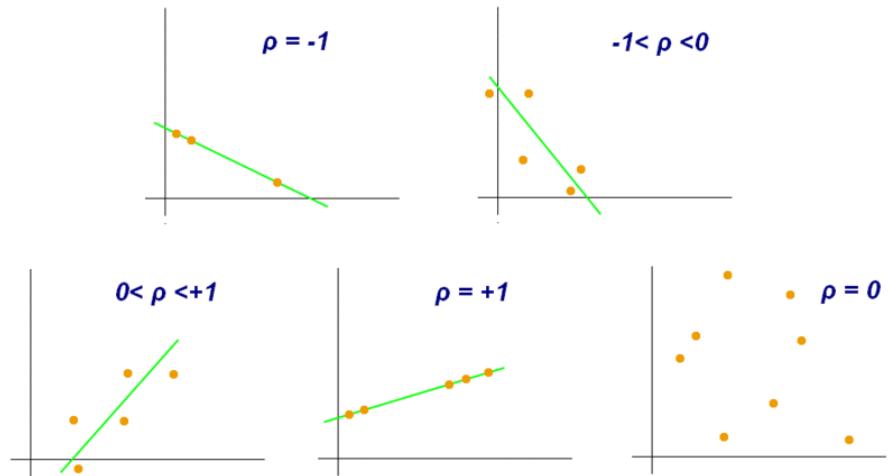


Individual metabolite level (2)

- **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:

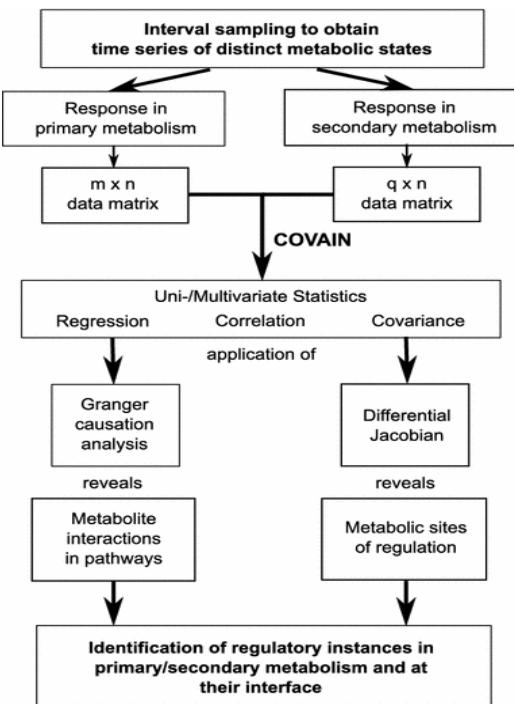
- 1) Correlation coefficients [0-1]
- 2) p-value to indicate significance



Individual metabolite level (3)

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

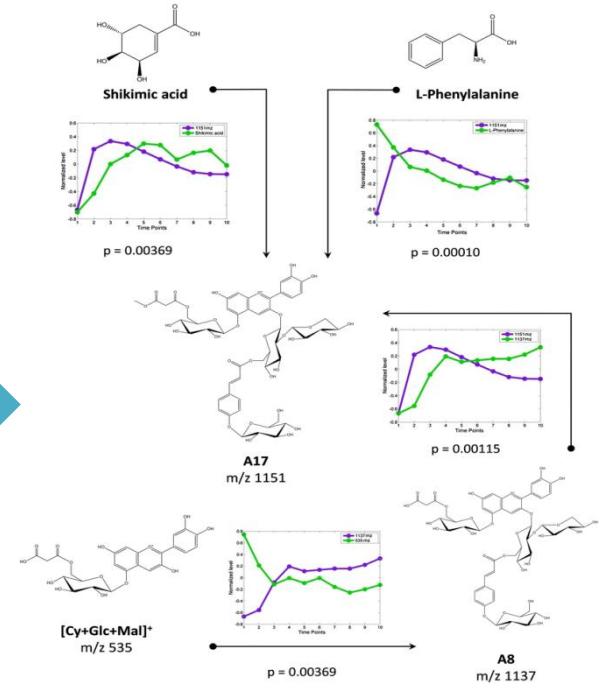
Hannes Doerfler*, David Lyon*, Xiaoliang Sun*, ..., Wolfram Weckwerth, Metabolomics, 2012



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

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

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

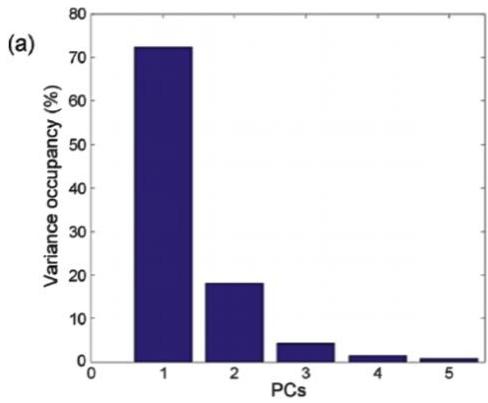


Metabolite group level (1)

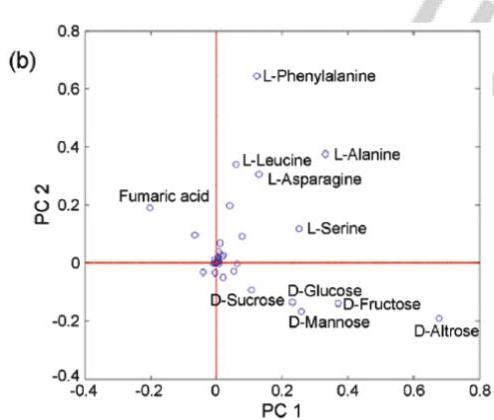
- Principal component analysis (PCA) distinguishes phenotypes and finds most influencing metabolites

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

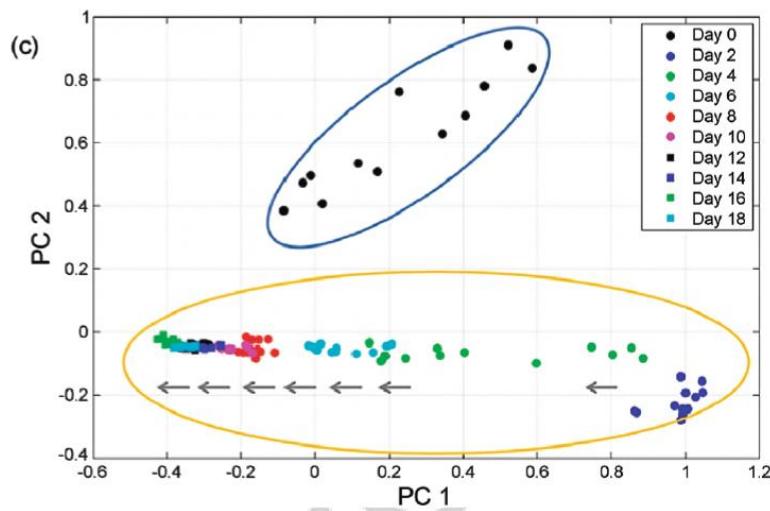
A. Variance occupancy



B. Loading plot



C. Coordinate plot

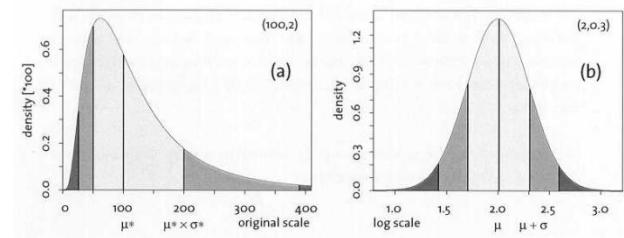
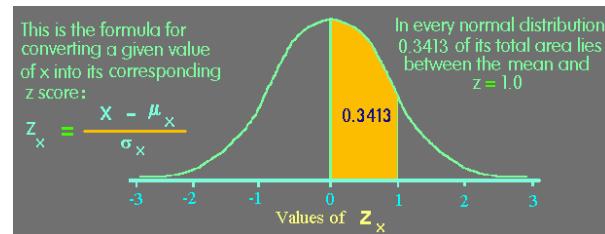


Data Scaling?

- Scaling is necessary if aiming to compare relative change

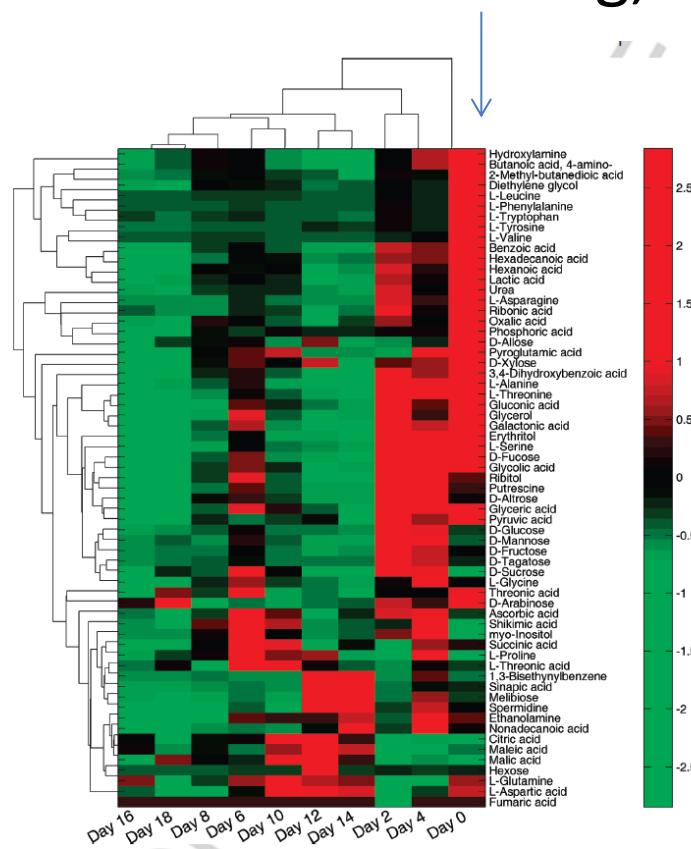
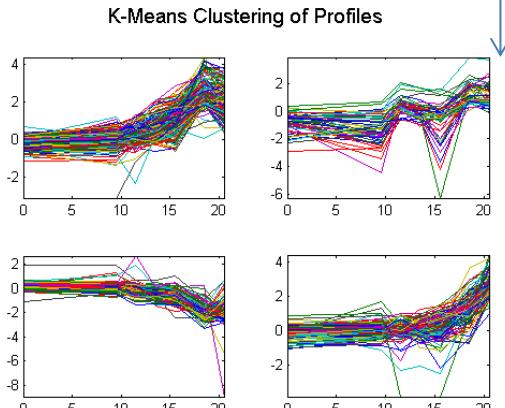
Metabolite	Initial Concentration [a.u.]	Fluctuation amplitude (%)	Changed ratio (fold)	Loading of PC1	Differential loading of PC1 (log2 ratio)
S ₁	1	10	10	0.0022	5.95
S ₂	10	10	5	0.0119	1.64
S ₃	100	10	2	0.0268	0.08
S ₄	1000	10	0 (no change)	0.9996	~ 0

- 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.4]
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-
Bonus	UVA+	UVB+	PAR+	N-
Bonus	UVA+	UVB+	PAR-	N-

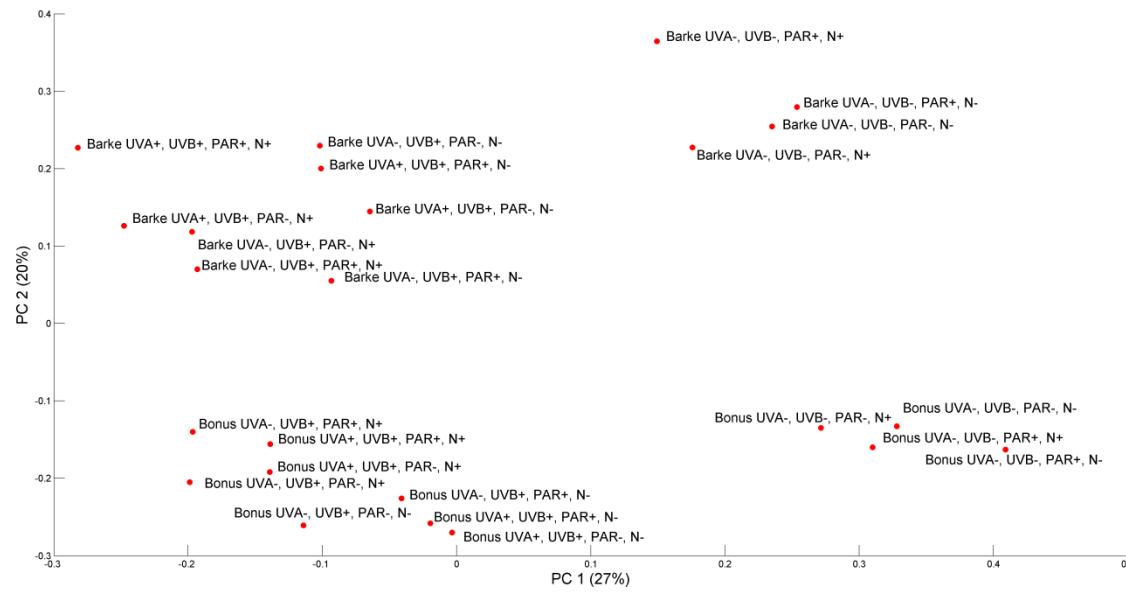
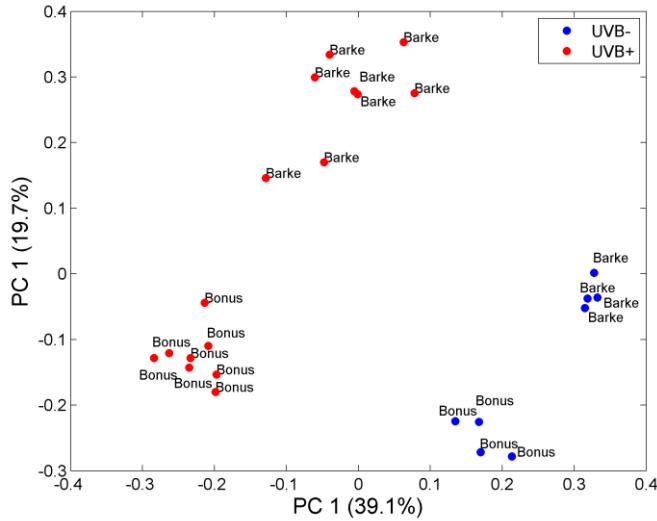
Discretize treatments: 1 for -, 2 for +

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

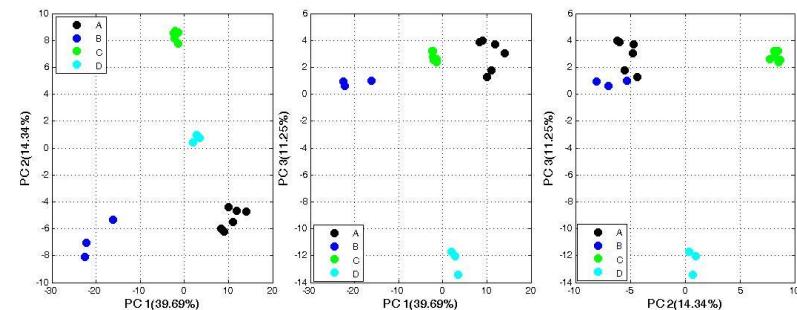
Statistical method	Metric	Missing value imputation & Outlier adjustment			Transformation & Scaling (column-wise)			
		Minimal	Mean	Distribution	Power	Log	Z-score	Range
ANOVA (t statistics)	Mean & Variance	●●	●	●	●	●	●	●
Correlation (Pearson's)	Mean	●●	○	●	●	●	○	○
Correlation (Spearman's)	Ranking of Mean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

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

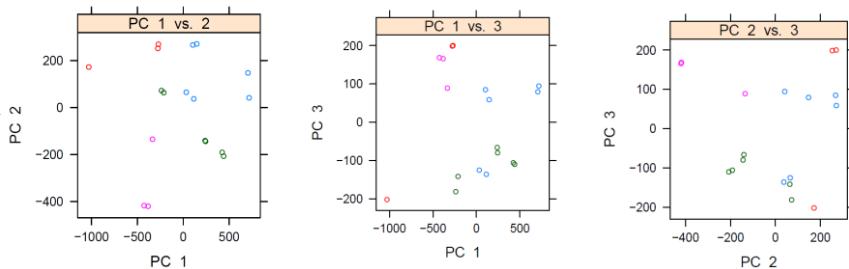
PCA analysis on a proteomics dataset

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

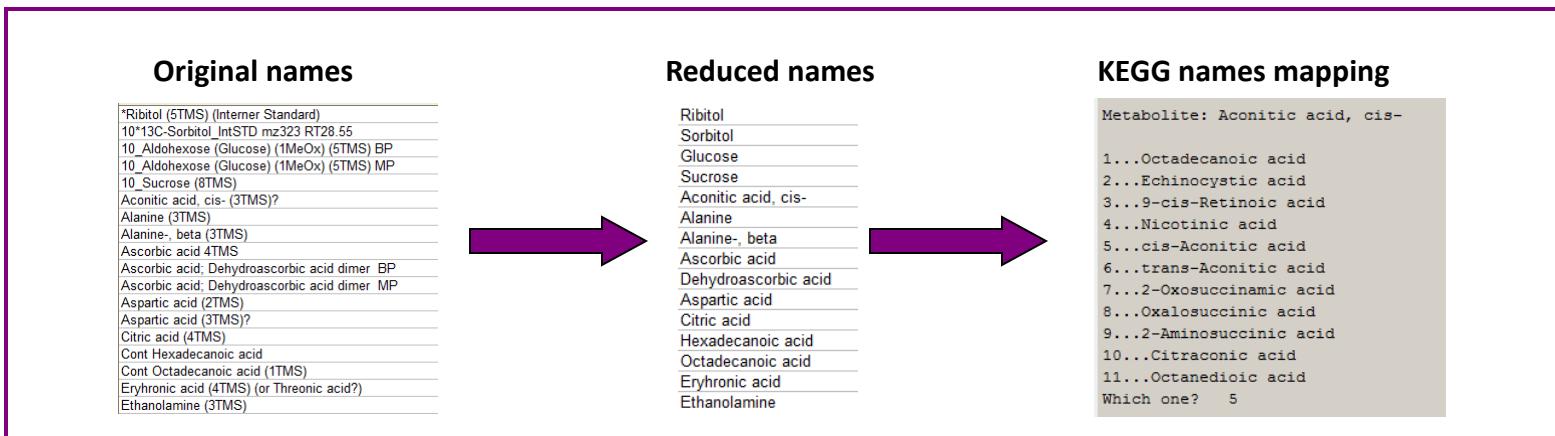
COVAIN results



MetaGeneAlyse results



Network level (1): database mapping



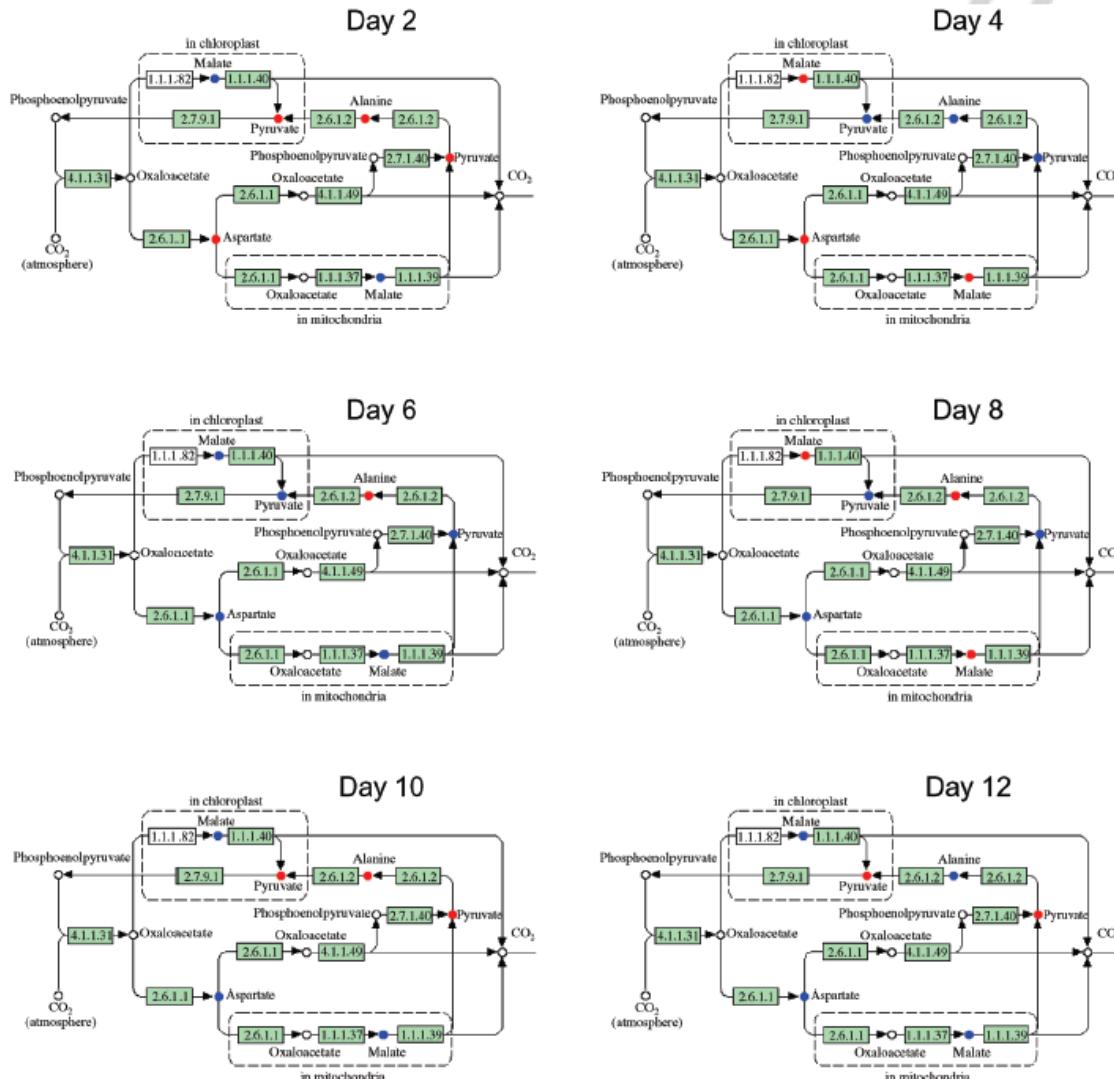
Name	KEGG Name	KEGG ID	Formula
Aconitic acid, cis-	cis-Aconitic acid	C00417	C6H6O6
Alanine	Alanine	C01401	C3H7NO2
Alanine-, beta	beta-Alanine	C00099	C3H7NO2
Ascorbic acid	Ascorbic acid	C00072	C6H8O6
Aspartic acid	Aspartic acid	C16433	C4H7NO4
Citric acid	Citric acid	C00158	C6H8O7
Dehydroascorbic acid	Dehydroascorbic acid	C05422	C6H6O6
Disaccharid	Lipid A disaccharide	C04932	C6H129N2O20P
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	C6H129N2O20P	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

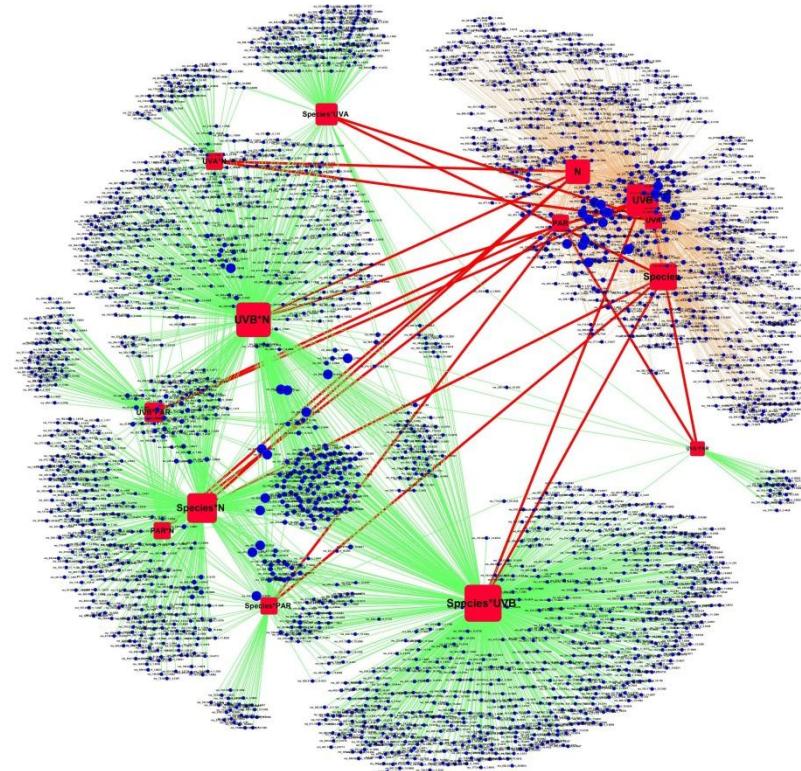
An example. Red: up- / Blue: down- regulated



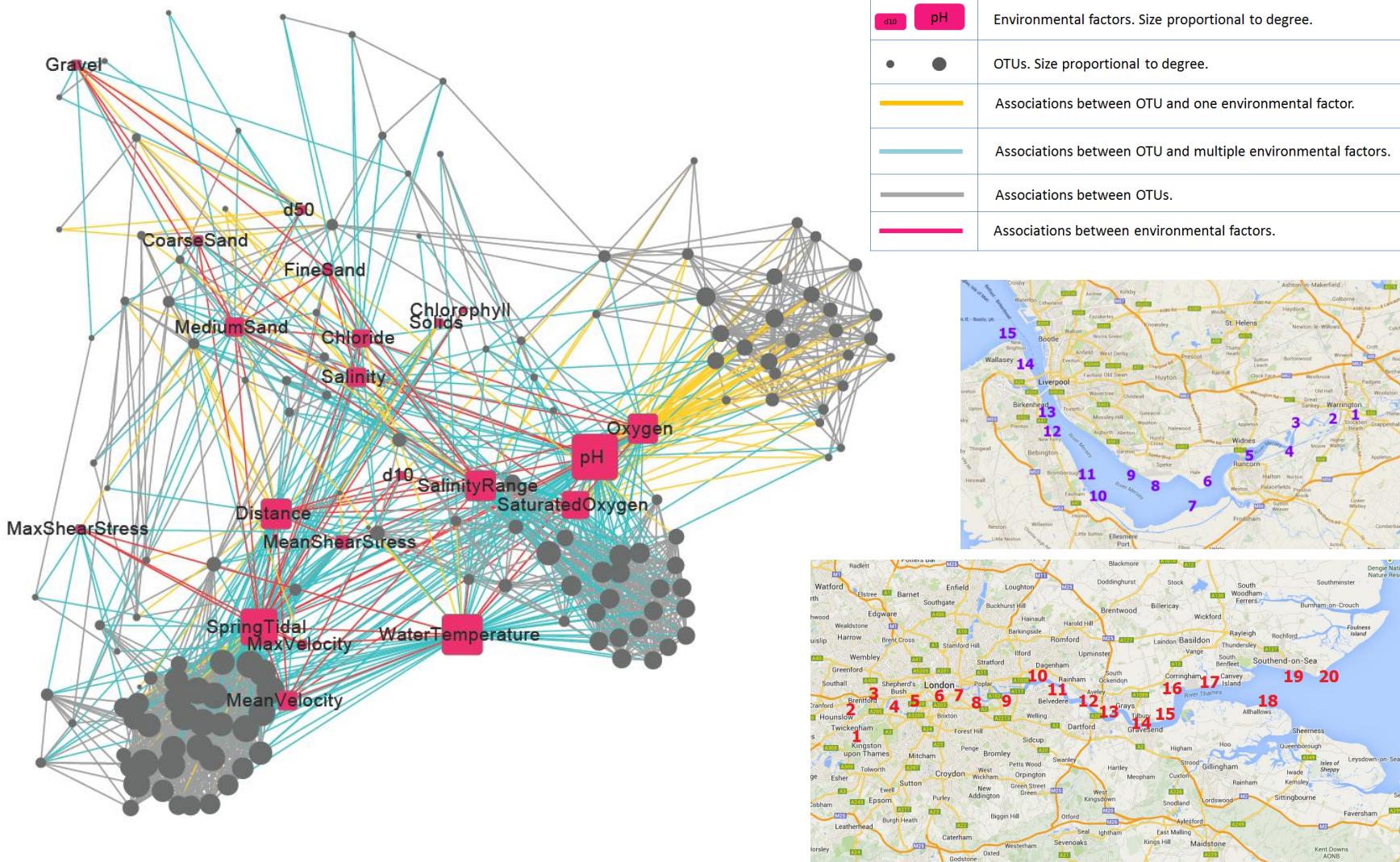
Pathway enrichment analysis can be done afterward.

Network level (2): inferring from data

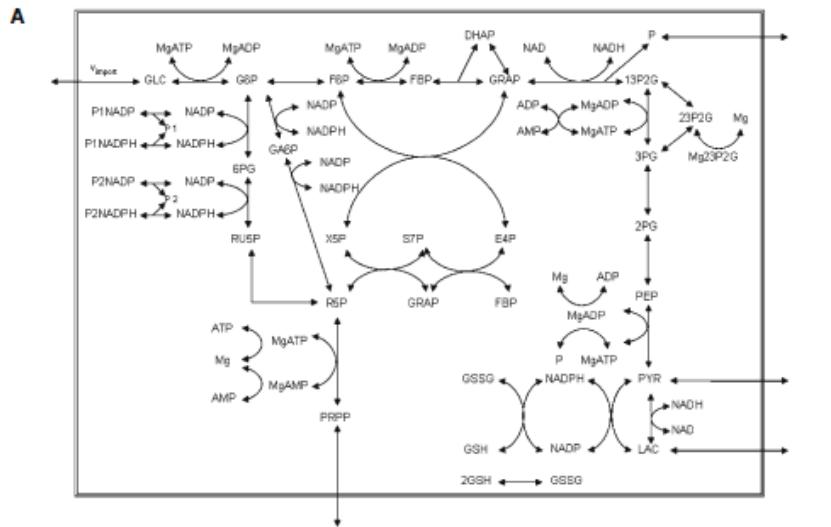
- Based on:
 - Correlation coefficients
 - Granger causality
 - Regression



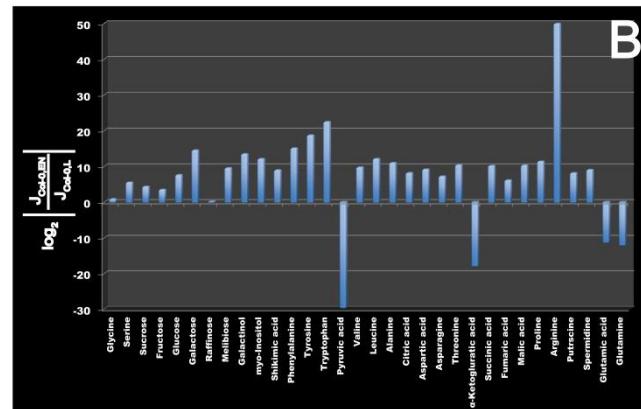
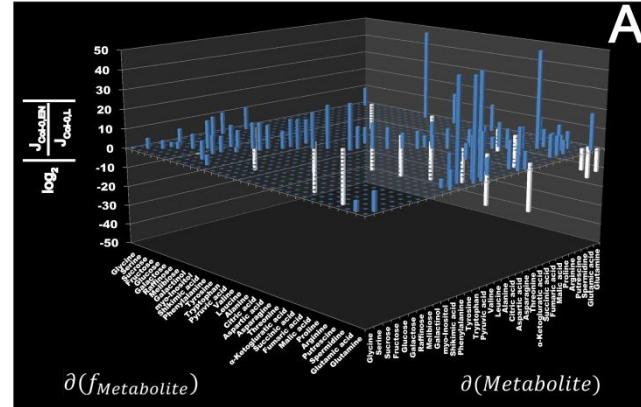
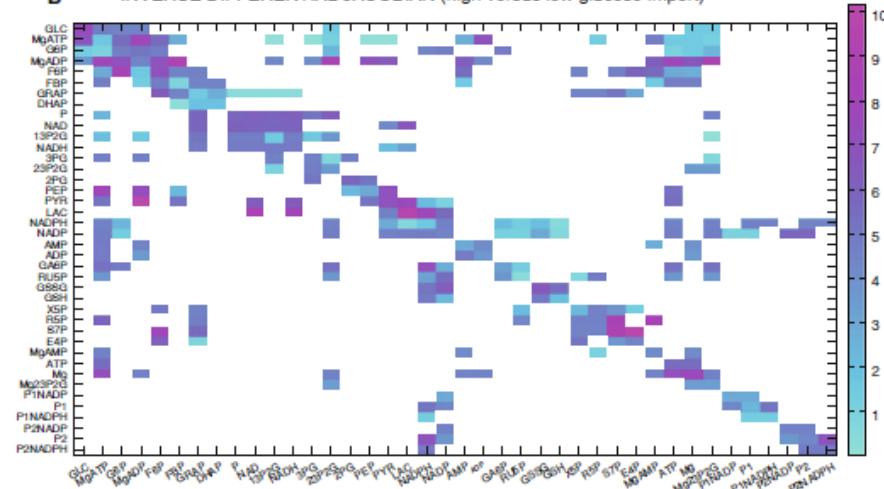
An ecological network



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

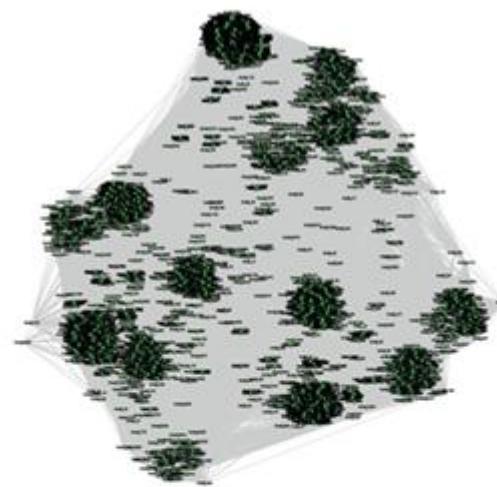
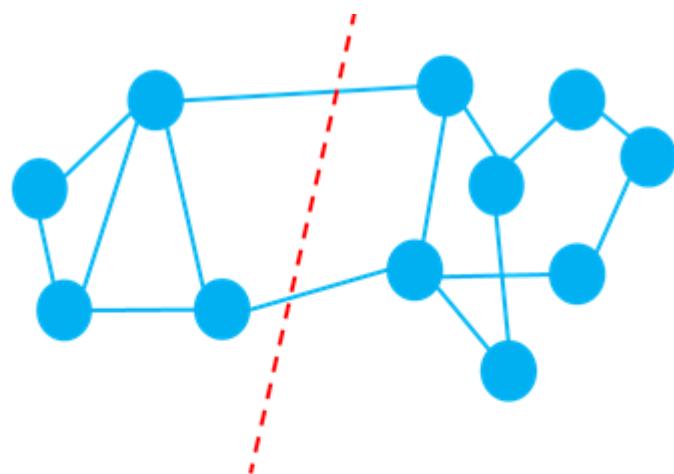


B INVERSE DIFFERENTIAL JACOBIAN (high versus low glucose import)

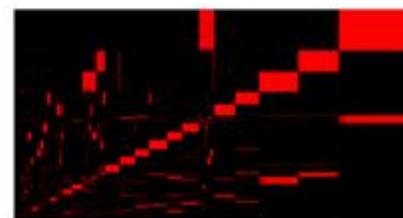


Sun X, Weckwerth W, Metabolomics, 2012
Nagele T, Mair A, Sun X, et al, PLoS One, 2014

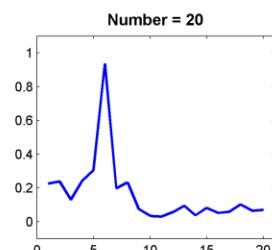
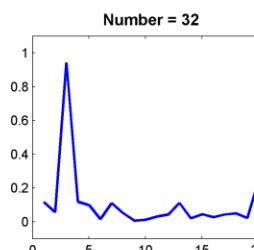
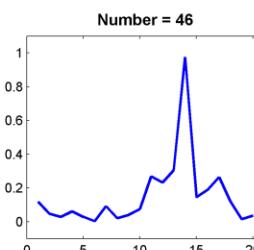
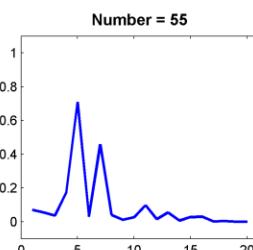
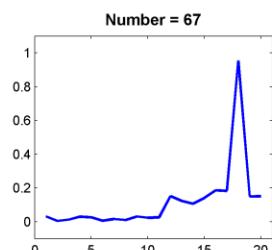
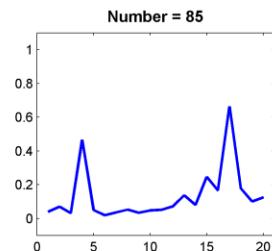
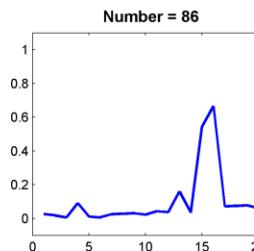
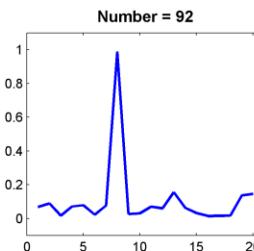
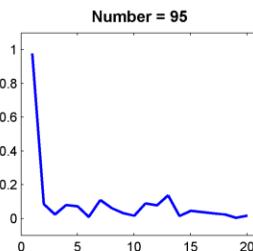
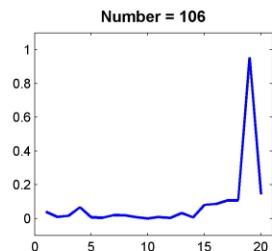
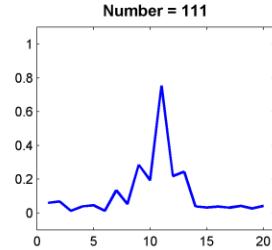
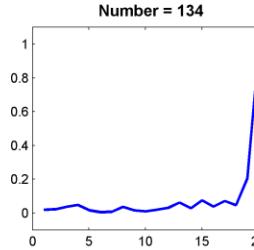
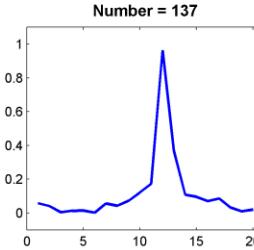
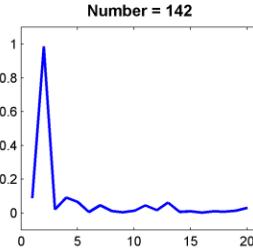
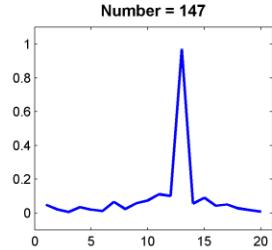
Network level (4): spectral clustering partitions network into separate modules



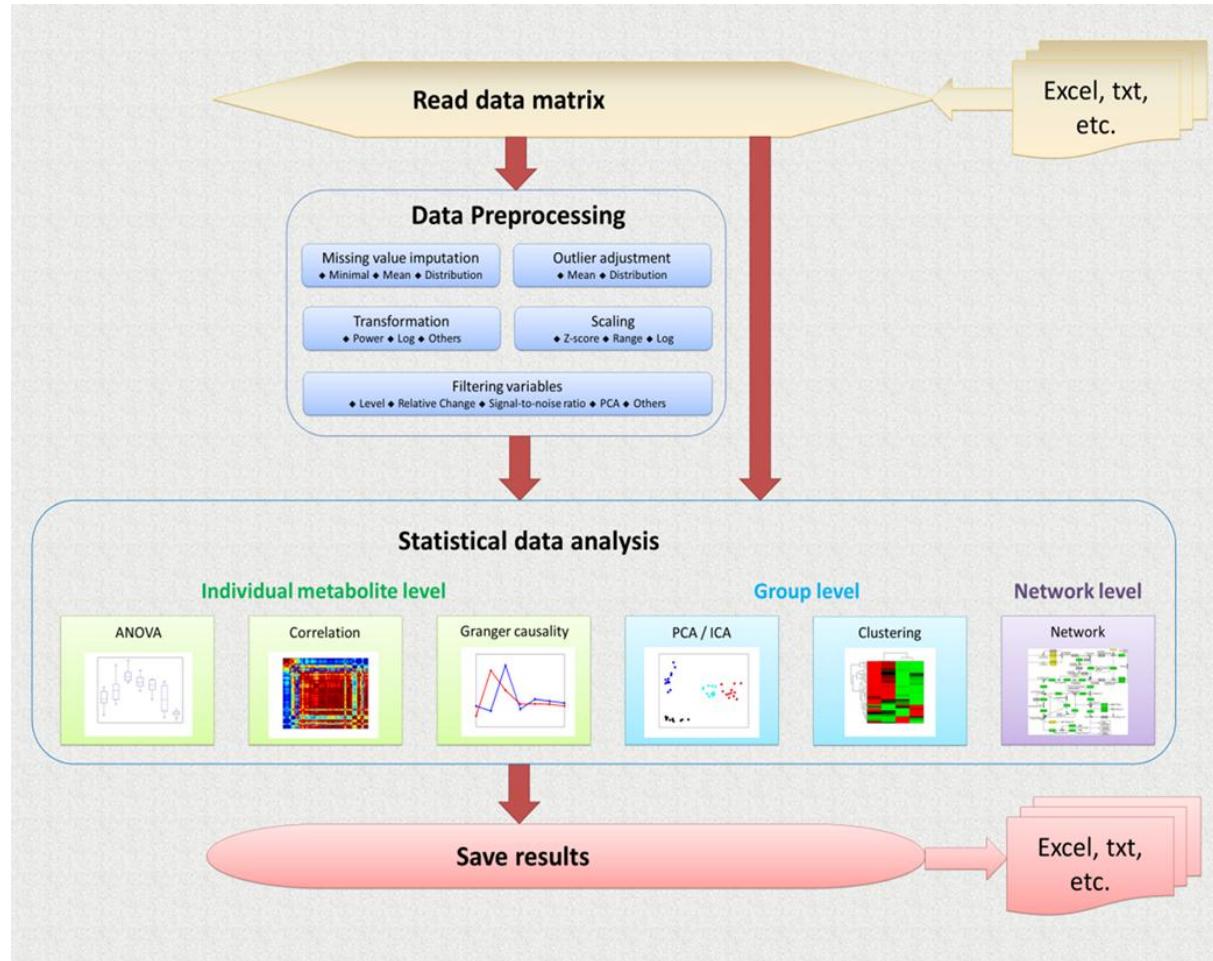
The network is constructed from regression coefficients and is partitioned into a few modules by spectral clustering.



Clustering variables from network



Analyzing metabolomics data by COVAIN



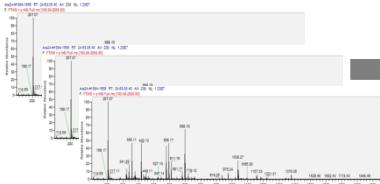
Sun X & Weckwerth W, Metabolomics, 2012
Doerfler H*, Sun X*, et al, PLoS One, 2014

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

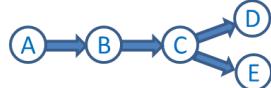
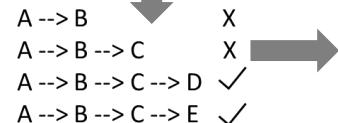
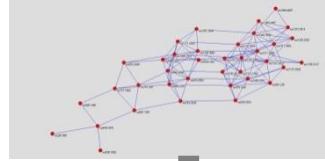
Read m/z data
and reaction
rules



A	B	C	D	E	F	Composition
448.11559	25978.6	0.86	448.11578	-0.41	C11 H22 O14 N5	
			448.11527	0.72	C25 H20 O8	
449.10783	188211.2	6.2	449.10783	-0.01	C20 H14 O3 N7	
			449.10784	-0.02	C11 H14 O11 N1	
594.18977	9240.6	0.3	594.18978	-0.85	C22 H28 O5 N4	
			594.18979	-0.87	C13 H24 O4 N2	
594.35642	8594.1	0.28	594.35644	-0.04	C37 H46 O3 N4	
			594.35695	-0.89	C23 H48 O9 N9	
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	

A	B	C	D	E
methylation / CH2 elongation... (+CH2)	1	2	0	0
methylation / H-methylation... (+CH3)	1	3	0	0
Demethylation / CH2 elongation... (-CH2)	2	4	0	0
Zermethylation / N-methylation... (+2xCH3)	2	6	0	0
monooxygenation (+O)	0	0	1	0
oxidation (+O2)	0	0	2	0
hydroxylation (+OH)	0	2	0	0
2xhydroxylation	0	4	0	0
X->X-cysteine cystein addition	3	5	2	1
X->X-kalanine addition	3	5	2	1
combined hydrogenation and protonation	0	3	0	0
protonation	0	1	0	0
hydrogenation	0	2	1	0
2xhydrogenation	0	4	2	0
oxidoreduction	0	-2	1	0
hexosylation (glycosylation...)	6	10	5	0
glycosylation (-O) / hexaminylation	6	10	4	0

mzGroupAnalyzer
+
Pathway Viewer



mzGroupAnalyzer Pathway Viewer

Select & List

Composition

From To

m/z value

Chemical Transformation Rules

From 700 to 1200

TP 1

Show Export

Totally 25 paths are found!

From (m/z2)	From (C14 H20 O24 N)	To (m/z)	TP 1
3: imz 728...	C14 H20 O24 N	imz 1198...	3: imz 990.2427...>
4: imz 728...	C45 H28 O8 N4	imz 1198...	4: imz 890.243...
5: imz 728...	C29 H30 O22 N	imz 1198...	5: imz 890.243...
6: imz 728...	C10 H16 O10 N2	imz 1198...	6: imz 890.243...
7: imz 757...	C22 H20 O5 N6	imz 1197...	7: imz 890.243...
8: imz 757...	C22 H20 O5 N6	imz 1197...	8: imz 890.243...
9: imz 887...	C40 H27 O17 N7	imz 1197...	9: imz 887...
10: imz 887...	C41 H43 O27 N7	imz 1197...	10: imz 887...
11: imz 887...	C50 H67 O37 N7	imz 1197...	11: imz 887...
12: imz 887...	C27 H45 O5 N5	imz 1197...	12: imz 887...
13: imz 978...	C24 H54 O32 N5	imz 1196...	13: imz 978...
14: imz 978...	C35 H64 O42 N6	imz 978...	14: imz 978...
15: imz 978...	C27 H45 O5 N5	imz 1196...	15: imz 978...
16: imz 978...	C35 H64 O42 N6	imz 978...	16: imz 978...
17: imz 978...	C27 H45 O5 N5	imz 1196...	17: imz 978...
18: imz 978...	C35 H64 O42 N6	imz 978...	18: imz 978...
19: imz 978...	C27 H45 O5 N5	imz 1196...	19: imz 978...
20: imz 978...	C35 H64 O42 N6	imz 978...	20: imz 978...
21: imz 978...	C27 H45 O5 N5	imz 1196...	21: imz 978...
22: imz 978...	C35 H64 O42 N6	imz 978...	22: imz 978...
23: imz 978...	C27 H45 O5 N5	imz 1196...	23: imz 978...
24: imz 978...	C35 H64 O42 N6	imz 978...	24: imz 978...
25: imz 978...	C27 H45 O5 N5	imz 1196...	25: imz 978...

Path:

Path 1: imz 887.2246 → imz 890.2427 → imz 1198.2985 → imz 1377.2994

File Edit View Insert Tools Desktop Window Help

hydrogenation (+H2) malonylation saponification monooxygenation (+O)

imz 887.2246 → imz 890.2427 → imz 875.2496 → imz 1181.2986 → imz 1197.2934

(C27 H45 O5 N5) (C27 H45 O28 N6) (C38 H45 O31 N6) (C41 H69 O35 N6) (C41 H69 O36 N6)

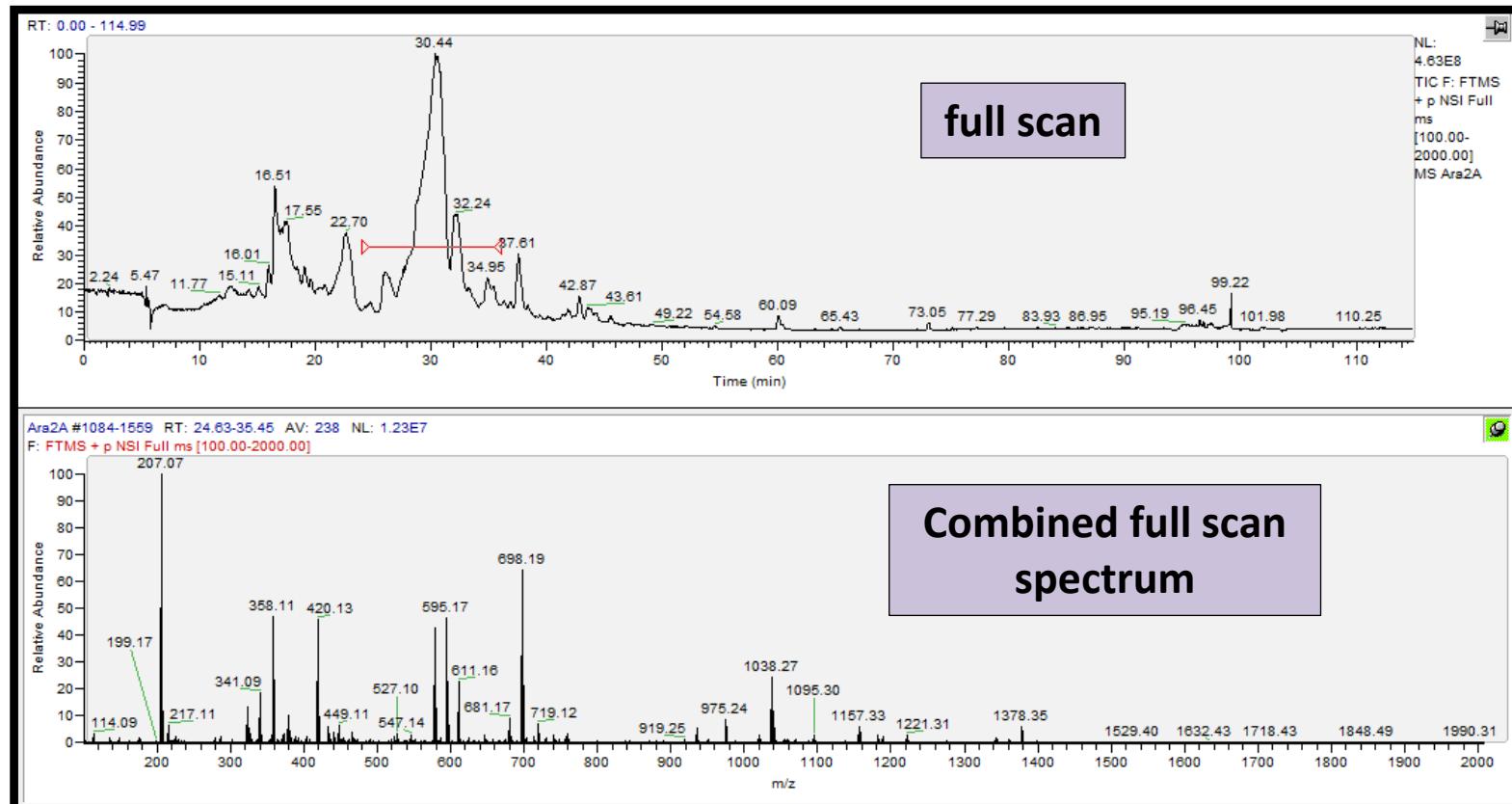
TP [1] TP [1] TP [1] TP [1]

Outputs

A	B	C	D	E	F
Pathway ID	Name	Reaction Rule	Pathway ID	Name	Reaction Rule
1	Pathway 1	Pathway 1	2	Pathway 2	Pathway 2
3	Pathway 3	Pathway 3	4	Pathway 4	Pathway 4
5	Pathway 5	Pathway 5	6	Pathway 6	Pathway 6
7	Pathway 7	Pathway 7	8	Pathway 8	Pathway 8
9	Pathway 9	Pathway 9	10	Pathway 10	Pathway 10
11	Pathway 11	Pathway 11	12	Pathway 12	Pathway 12
13	Pathway 13	Pathway 13	14	Pathway 14	Pathway 14
15	Pathway 15	Pathway 15	16	Pathway 16	Pathway 16
17	Pathway 17	Pathway 17	18	Pathway 18	Pathway 18
19	Pathway 19	Pathway 19	20	Pathway 20	Pathway 20
21	Pathway 21	Pathway 21	22	Pathway 22	Pathway 22
23	Pathway 23	Pathway 23	24	Pathway 24	Pathway 24
25	Pathway 25	Pathway 25	26	Pathway 26	Pathway 26
27	Pathway 27	Pathway 27	28	Pathway 28	Pathway 28
29	Pathway 29	Pathway 29	30	Pathway 30	Pathway 30
31	Pathway 31	Pathway 31	32	Pathway 32	Pathway 32
33	Pathway 33	Pathway 33	34	Pathway 34	Pathway 34
35	Pathway 35	Pathway 35	36	Pathway 36	Pathway 36
37	Pathway 37	Pathway 37	38	Pathway 38	Pathway 38
39	Pathway 39	Pathway 39	40	Pathway 40	Pathway 40
41	Pathway 41	Pathway 41	42	Pathway 42	Pathway 42
43	Pathway 43	Pathway 43	44	Pathway 44	Pathway 44
45	Pathway 45	Pathway 45	46	Pathway 46	Pathway 46
47	Pathway 47	Pathway 47	48	Pathway 48	Pathway 48
49	Pathway 49	Pathway 49	50	Pathway 50	Pathway 50
51	Pathway 51	Pathway 51	52	Pathway 52	Pathway 52
53	Pathway 53	Pathway 53	54	Pathway 54	Pathway 54
55	Pathway 55	Pathway 55	56	Pathway 56	Pathway 56
57	Pathway 57	Pathway 57	58	Pathway 58	Pathway 58
59	Pathway 59	Pathway 59	60	Pathway 60	Pathway 60
61	Pathway 61	Pathway 61	62	Pathway 62	Pathway 62
63	Pathway 63	Pathway 63	64	Pathway 64	Pathway 64
65	Pathway 65	Pathway 65	66	Pathway 66	Pathway 66
67	Pathway 67	Pathway 67	68	Pathway 68	Pathway 68
69	Pathway 69	Pathway 69	70	Pathway 70	Pathway 70
71	Pathway 71	Pathway 71	72	Pathway 72	Pathway 72
73	Pathway 73	Pathway 73	74	Pathway 74	Pathway 74
75	Pathway 75	Pathway 75	76	Pathway 76	Pathway 76
77	Pathway 77	Pathway 77	78	Pathway 78	Pathway 78
79	Pathway 79	Pathway 79	80	Pathway 80	Pathway 80
81	Pathway 81	Pathway 81	82	Pathway 82	Pathway 82
83	Pathway 83	Pathway 83	84	Pathway 84	Pathway 84
85	Pathway 85	Pathway 85	86	Pathway 86	Pathway 86
87	Pathway 87	Pathway 87	88	Pathway 88	Pathway 88
89	Pathway 89	Pathway 89	90	Pathway 90	Pathway 90
91	Pathway 91	Pathway 91	92	Pathway 92	Pathway 92
93	Pathway 93	Pathway 93	94	Pathway 94	Pathway 94
95	Pathway 95	Pathway 95	96	Pathway 96	Pathway 96
97	Pathway 97	Pathway 97	98	Pathway 98	Pathway 98
99	Pathway 99	Pathway 99	100	Pathway 100	Pathway 100
101	Pathway 101	Pathway 101	102	Pathway 102	Pathway 102
103	Pathway 103	Pathway 103	104	Pathway 104	Pathway 104
105	Pathway 105	Pathway 105	106	Pathway 106	Pathway 106
107	Pathway 107	Pathway 107	108	Pathway 108	Pathway 108
109	Pathway 109	Pathway 109	110	Pathway 110	Pathway 110
111	Pathway 111	Pathway 111	112	Pathway 112	Pathway 112
113	Pathway 113	Pathway 113	114	Pathway 114	Pathway 114
115	Pathway 115	Pathway 115	116	Pathway 116	Pathway 116
117	Pathway 117	Pathway 117	118	Pathway 118	Pathway 118
119	Pathway 119	Pathway 119	120	Pathway 120	Pathway 120
121	Pathway 121	Pathway 121	122	Pathway 122	Pathway 122
123	Pathway 123	Pathway 123	124	Pathway 124	Pathway 124
125	Pathway 125	Pathway 125	126	Pathway 126	Pathway 126
127	Pathway 127	Pathway 127	128	Pathway 128	Pathway 128
129	Pathway 129	Pathway 129	130	Pathway 130	Pathway 130
131	Pathway 131	Pathway 131	132	Pathway 132	Pathway 132
133	Pathway 133	Pathway 133	134	Pathway 134	Pathway 134
135	Pathway 135	Pathway 135	136	Pathway 136	Pathway 136
137	Pathway 137	Pathway 137	138	Pathway 138	Pathway 138
139	Pathway 139	Pathway 139	140	Pathway 140	Pathway 140
141	Pathway 141	Pathway 141	142	Pathway 142	Pathway 142
143	Pathway 143	Pathway 143	144	Pathway 144	Pathway 144
145	Pathway 145	Pathway 145	146	Pathway 146	Pathway 146
147	Pathway 147	Pathway 147	148	Pathway 148	Pathway 148
149	Pathway 149	Pathway 149	150	Pathway 150	Pathway 150
151	Pathway 151	Pathway 151	152	Pathway 152	Pathway 152
153	Pathway 153	Pathway 153	154	Pathway 154	Pathway 154
155	Pathway 155	Pathway 155	156	Pathway 156	Pathway 156
157	Pathway 157	Pathway 157	158	Pathway 158	Pathway 158
159	Pathway 159	Pathway 159	160	Pathway 160	Pathway 160
161	Pathway 161	Pathway 161	162	Pathway 162	Pathway 162
163	Pathway 163	Pathway 163	164	Pathway 164	Pathway 164
165	Pathway 165	Pathway 165	166	Pathway 166	Pathway 166
167	Pathway 167	Pathway 167	168	Pathway 168	Pathway 168
169	Pathway 169	Pathway 169	170	Pathway 170	Pathway 170
171	Pathway 171	Pathway 171	172	Pathway 172	Pathway 172
173	Pathway 173	Pathway 173	174	Pathway 174	Pathway 174
175	Pathway 175	Pathway 175	176	Pathway 176	Pathway 176
177	Pathway 177	Pathway 177	178	Pathway 178	Pathway 178
179	Pathway 179	Pathway 179	180	Pathway 180	Pathway 180
181	Pathway 181	Pathway 181	182	Pathway 182	Pathway 182
183	Pathway 183	Pathway 183	184	Pathway 184	Pathway 184
185	Pathway 185	Pathway 185	186	Pathway 186	Pathway 186
187	Pathway 187	Pathway 187	188	Pathway 188	Pathway 188
189	Pathway 189	Pathway 189	190	Pathway 190	Pathway 190
191	Pathway 191	Pathway 191	192	Pathway 192	Pathway 192
193	Pathway 193	Pathway 193	194	Pathway 194	Pathway 194
195	Pathway 195	Pathway 195	196	Pathway 196	Pathway 196
197	Pathway 197	Pathway 197	198	Pathway 198	Pathway 198
199	Pathway 199	Pathway 199	200	Pathway 200	Pathway 200
201	Pathway 201	Pathway 201	202	Pathway 202	Pathway 202
203	Pathway 203	Pathway 203	204	Pathway 204	Pathway 204
205	Pathway 205	Pathway 205	206	Pathway 206	Pathway 206
207	Pathway 207	Pathway 207	208	Pathway 208	Pathway 208
209	Pathway 209	Pathway 209	210	Pathway 210	Pathway 210
211	Pathway 211	Pathway 211	212	Pathway 212	Pathway 212
213	Pathway 213	Pathway 213	214	Pathway 214	Pathway 214
215	Pathway 215	Pathway 215	216	Pathway 216	Pathway 216
217	Pathway 217	Pathway 217	218	Pathway 218	Pathway 218
219	Pathway 219	Pathway 219	220	Pathway 220	Pathway 220
221	Pathway 221	Pathway 221	222	Pathway 222	Pathway 222
223	Pathway 223	Pathway 223	224	Pathway 224	Pathway 224
225	Pathway 225	Pathway 225	226	Pathway 226	Pathway 226
227	Pathway 227	Pathway 227	228	Pathway 228	Pathway 228
229	Pathway 229	Pathway 229	230	Pathway 230	Pathway 230
231	Pathway 231	Pathway 231	232	Pathway 232	Pathway 232
233	Pathway 233	Pathway 233	234	Pathway 234	Pathway 234
235	Pathway 235	Pathway 235	236	Pathway 236	Pathway 236
237	Pathway 237	Pathway 237	238	Pathway 238	Pathway 238
239	Pathway 239	Pathway 239	240	Pathway 240	Pathway 240
241	Pathway 241	Pathway 241	242	Pathway 242	Pathway 242
243	Pathway 243	Pathway 243	244	Pathway 244	Pathway 244
245	Pathway 245	Pathway 245	246	Pathway 246	Pathway 246
247	Pathway 247	Pathway 247	248	Pathway 248	Pathway 248
249	Pathway 249	Pathway 249			

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
- ...

Important:

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

RT: 0.00 - 114.99

Time (min)

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

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

F: PFMMS + n NST Full ms [100.00-2000.001]

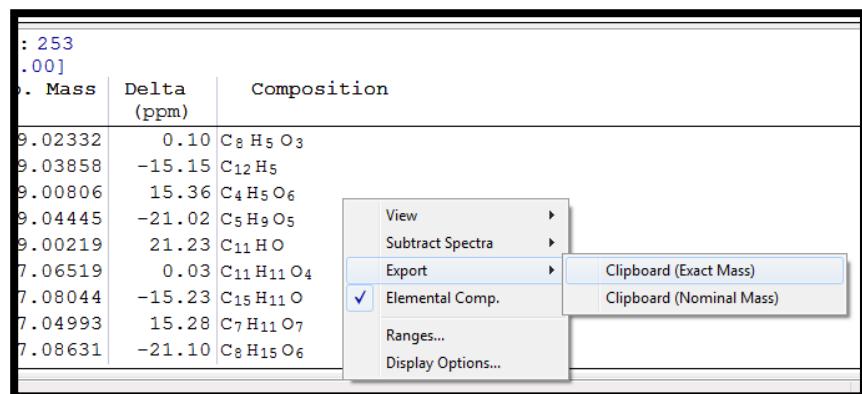
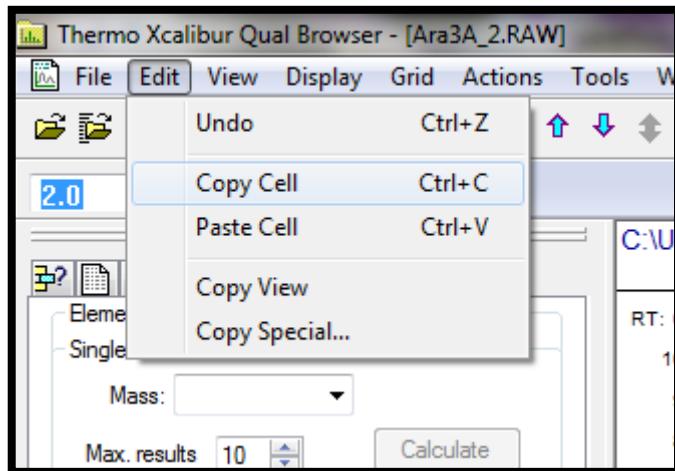
m/z	Intensity	Relative	Theo. Mass	Delta (ppm)	Composition
100.07584	181551.6	3.24	100.07569	0.15	C ₅ H ₁₀ O N
114.09143	403032.5	7.20	114.09134	0.09	C ₆ H ₁₂ O N
135.10163	261628.9	4.67	135.10157	0.06	C ₆ H ₁₅ O ₃
136.02160	240385.4	4.29	136.02271	-1.11	C H ₄ O ₄ N ₄
147.04411	122086.8	2.18	147.04406	0.06	C ₉ H ₇ O ₂
149.02336	236602.6	4.23	149.02332	0.04	C ₈ H ₅ O ₃
163.13291	142901.4	2.55	163.13287	0.04	C ₈ H ₁₉ O ₃
177.05465	155743.0	2.78	177.05462	0.03	C ₁₀ H ₉ O ₃
193.04954	379139.6	6.77	193.04954	0.01	C ₁₀ H ₉ O ₄

3. Export from Xcalibur to Excel

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

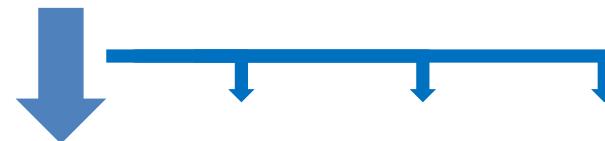
alternatively

- „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

Data information

Choose data

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

Transpose

Normalize

Fill missing values

Adjust outliers

ANOVA

Plot (all)

Evaluation

Log transform

Z transform

0-1 Range scaling

COVAIN guided user interface

Load data ...

Combine data ...

My notes ...

Name the results

Save

Options

Help

Data Analysis

Multivariate Statistics

Data selection

Input index, such as 1,2,3 All

PCA

PC number (max 5) Plot

ICA

IC number (max 5) Plot

Correlation

Heatmap

Cluster

Normal analysis Case-Control analysis

PCA Label

ICA Label

Time Series

Time points selection

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

Correlation

Calculate & View Heatmap

Clustering

Hierarchical K-Means

Cluster

Granger causation analysis

Calculate & View

Permutation Entropy

Calculate & View

Network Analysis

Network Inference

Correlation Correlation (Time Series) Visualize network

Network Property

Calculate & View

Inverse Jacobian

Set conditions

Set conditions

KEGG Pathway

Compound mapping ...

mz GroupAnalyzer

Load data & Analyze Pathway Viewer

1. load data

2. select data, rule file and output location

A	B	C	D	E	F	G
21	595.35042	8584.1	0.28	256.25564	-0.04	C37 H44 O9 N3 N2
22			594.25695	-0.39	C34 H45 O9 N2	
23			594.25996	-0.39	C34 H45 O9 N2	
24	595.34474	71225.1	2.35	232.25564	-0.04	C37 H44 O9 N3 N2
25			595.34463	-0.22	C29 H21 O8 N7	
26			595.44512	-0.64	C16 H29 O13 N5	
27	595.16563	841436.1	27.73	232.25564	-0.04	C37 H44 O9 N3 N2
28			595.34875	-0.19	C27 H21 O15	
29			595.36323	-0.67	C40 H32 O14 N2	
30	595.36888	32341.7	0.41	595.36888	0.00	C37 H44 O9 N3 N2
31			595.36888	-0.01	C29 H21 O8 N7	
32			595.36882	0.85	C42 H47 O12 N2	
33	727.18993	20668.8	0.68	727.18868	0.08	C10 H13 O17
34			727.18993	0.08	C10 H13 O17	
35			727.18718	-0.62	C21 H37 O21 N5	
36			727.18718	-0.78	C48 H72 O12 N2	
37			727.18718	0.78	C48 H72 O12 N2	
38	727.35777	11794	0.45	727.35756	-0.28	C40 H49 O8 N5
39			727.35756	-0.42	C36 H51 O14 N10	
40			727.35756	0.00	C36 H51 O14 N10	
41			727.35756	0.97	C44 H47 O7	
42	728.19048	7156.8	0.24	728.19067	-0.26	C29 H30 O14 N9
43			728.19067	0.00	C29 H30 O14 N9	
44			728.19017	0.43	C43 H52 O8 N4	
45			728.19084	0.87	C14 H38 O24 N10	

data file

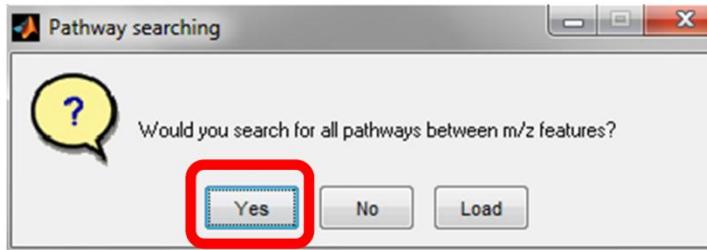
A	B	C	D	E
1	methionine (C) / chemical modification	methionine (C) / chemical modification	1	0
2	methylation / N-methylation... (-CH3)		1	3
3	methylation / N-methylation... (-CH3)		28.0113001	0
4	2-methylation / C/H elongation... (+2xCH2)		30.0100002	0
5	2-methylation / C/H elongation... (+2xCH2)		35.9985146	0
6	monoxygenation (+O)		31.9890293	0
7	oxidation (+O2)		1.0100007	0
8	polymerization (+CH2)		4.63130013	0
9	2hydrogenation		1.0100007	0
10	combined hydrogenation and protonation		1.0100007	0
11	dehydration		18.0105647	0
12	hydration		36.0211294	0
13	dehydration		13.9730048	0
14	hydration		16.0250243	1
15	ketoylation (glycation...)		162.052024	6
16	glycation (C) / chymotrypsin		144.05799	6
17	dehydration		138.002094	3
18	pentamethylation (ribosylation...)		132.042239	5
19	malonylation		88.000394	3
20	carboxylation		184.00000	2
21	decarboxylation		208.05799	11
22	isopropylation		43.9890293	1
23	carboxylation		204.000233	0
24	acetylation		42.0105647	2
25	etherification		30.0105647	1
26	(A) methylation (B) e-amino acid formation from orn C) C=OH "insertion" after			

rule file

+ output directory

→ transformation files will be created

3. start Pathway Viewer



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



Select & List

Composition

C H O

From To

m / z value

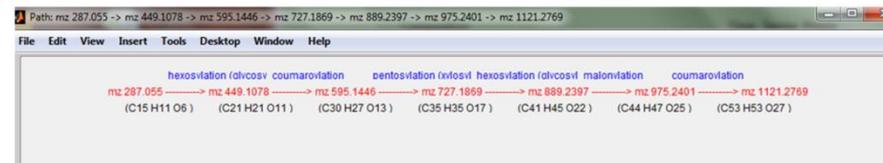
From to

Chemical Transformation Rules

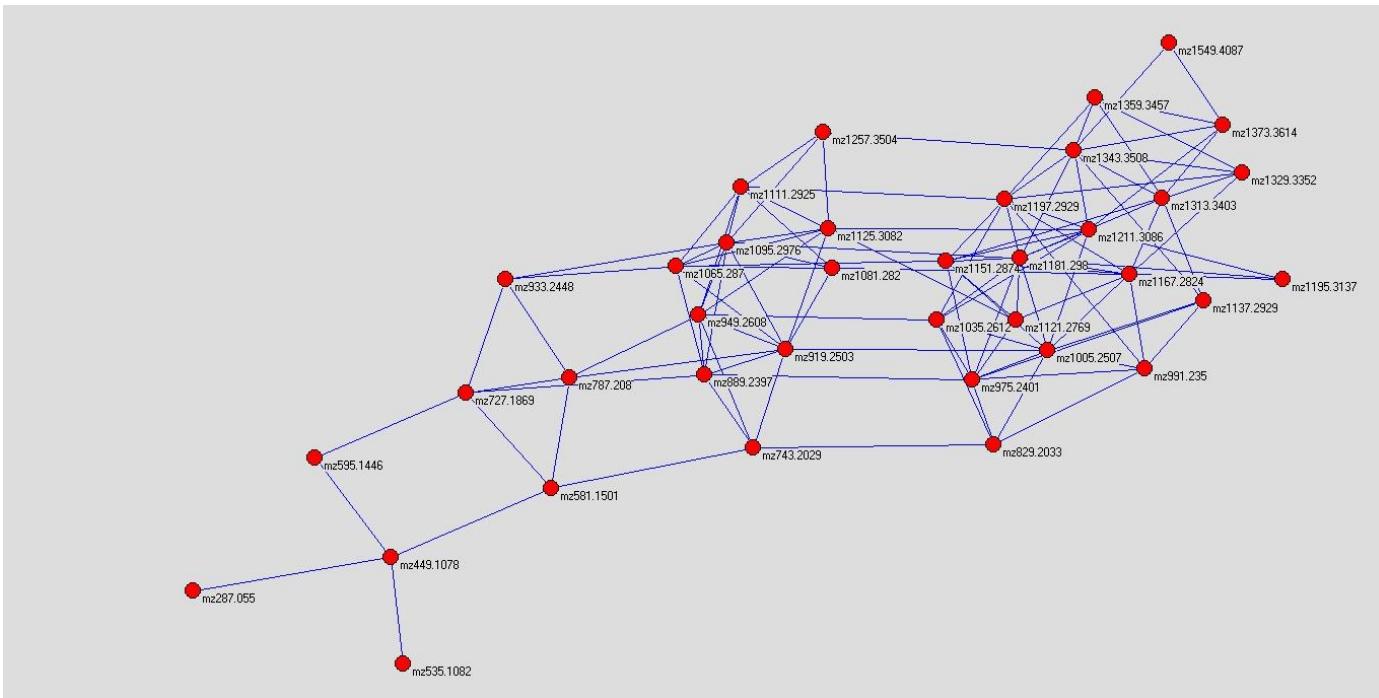
OR

	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
3	mz 287.055	C15 H11 O6	mz 1005.2...	C45 H49 O...	mz 287.055 -> mz 449.1078 -> mz 581.1501 -> mz
4	mz 287.055	C15 H11 O6	mz 1035.2...	C46 H51 O...	mz 287.055 -> mz 449.1078 -> mz 581.1501 -> mz
5	mz 287.055	C15 H11 O6	mz 1065.287	C51 H53 O...	mz 287.055 -> mz 449.1078 -> mz 595.1446 -> mz
6	mz 287.055	C15 H11 O6	mz 1081.282	C51 H53 O...	mz 287.055 -> mz 449.1078 -> mz 581.1501 -> mz
7	mz 287.055	C15 H11 O6	mz 1111.2...	C52 H55 O...	mz 287.055 -> mz 449.1078 -> mz 581.1501 -> mz
8	mz 287.055	C15 H11 O6	mz 1121.2...	C53 H53 O...	mz 287.055 -> mz 449.1078 -> mz 581.1501 -> mz
9	mz 287.055	C15 H11 O6	mz 1151.2...	C54 H55 O...	mz 287.055 -> mz 449.1078 -> mz 581.1501 -> mz
10	mz 287.055	C15 H11 O6	mz 1167.2...	C54 H55 O...	mz 287.055 -> mz 449.1078 -> mz 581.1501 -> mz
11	mz 287.055	C15 H11 O6	mz 1195.3...	C56 H59 O...	mz 287.055 -> mz 449.1078 -> mz 581.1501 -> mz
12	mz 287.055	C15 H11 O6	mz 1257.3...	C58 H65 O...	mz 287.055 -> mz 449.1078 -> mz 581.1501 -> mz
13	mz 287.055	C15 H11 O6	mz 1313.3...	C60 H65 O...	mz 287.055 -> mz 449.1078 -> mz 581.1501 -> mz
14	mz 287.055	C15 H11 O6	mz 1329.3...	C60 H65 O...	mz 287.055 -> mz 449.1078 -> mz 581.1501 -> mz
15	mz 287.055	C15 H11 O6	mz 1359.3...	C61 H67 O...	mz 287.055 -> mz 449.1078 -> mz 581.1501 -> mz
16	mz 287.055	C15 H11 O6	mz 1373.3...	C62 H69 O...	mz 287.055 -> mz 449.1078 -> mz 581.1501 -> mz
17	mz 287.055	C15 H11 O6	mz 1549.4...	C72 H77 O...	mz 287.055 -> mz 449.1078 -> mz 581.1501 -> mz
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 -> n
20	mz 595.1446	C30 H27 O...	mz 949.2608	C41 H45 O...	mz 595.1446 -> mz 727.1869 -> mz 787.208 -> mz
21	mz 595.1446	C30 H27 O...	mz 991.208	C44 H47 O...	mz 595.1446 -> mz 727.1869 -> mz 889.2397 -> n
22	mz 595.1446	C30 H27 O...	mz 1005.2...	C45 H49 O...	mz 595.1446 -> mz 727.1869 -> mz 889.2397 -> n

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.