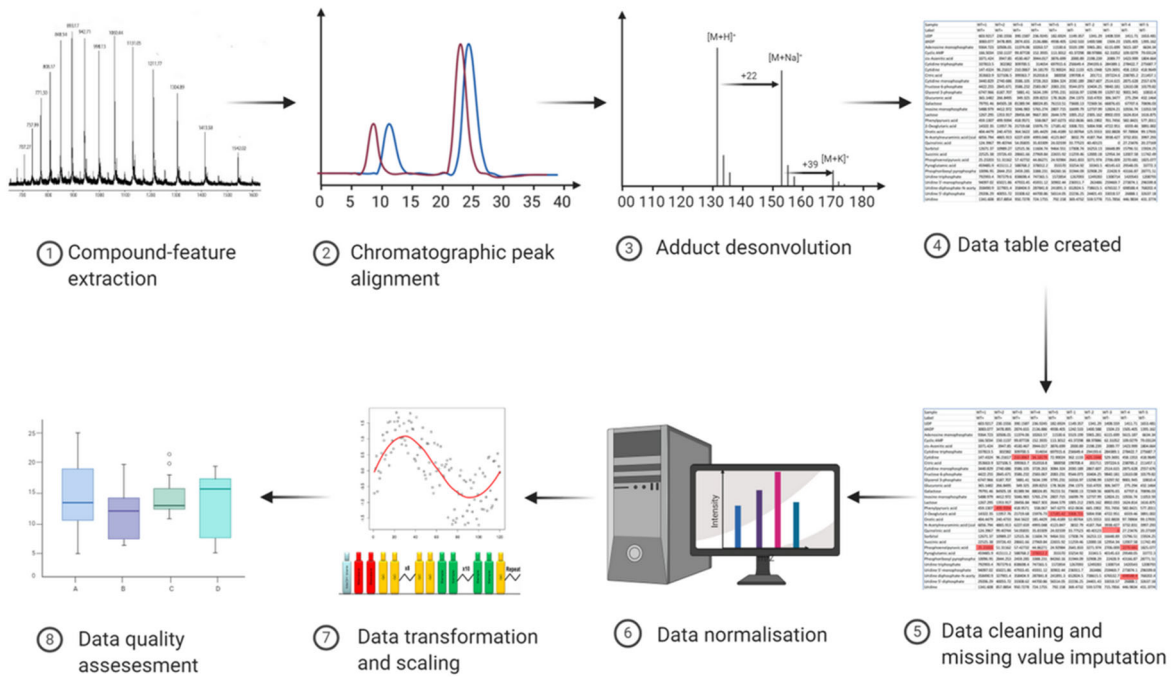


Lecture 1 Handout



Metabolomics Data Processing and Statistical Analysis

OVERVIEW

Part 1: Introduction to processing and analysis of metabolomics data

- The purpose of bioinformatics in metabolomics
- Bioinformatics workflow
- Pre-processing data (raw files to data tables)
- Final data processing data (data formatting for statistical analysis)
- Statistical tools: univariate
- Statistical tools: multivariate

Part 2: MetaboAnalyst – data processing and statistical analysis

- Introduction to MetaboAnalyst software
- Data processing (preparing data for statistical analysis)
- Statistical analysis: fold-change, significance, supervised and unsupervised multivariate analysis, correlations and classification.
- Data reporting

DEFINING METABOLOMICS

Metabolomics as an analytical approach which aims to **comprehensively characterise metabolites in a biological system in order to investigate altered metabolic states**.

Targeted and untargeted approaches can be used which provide **hypothesis generating** and **hypothesis-led** experiments leading to complementary information that contributes to a systems-level understanding of the biological context and subsequently a molecular phenotype.

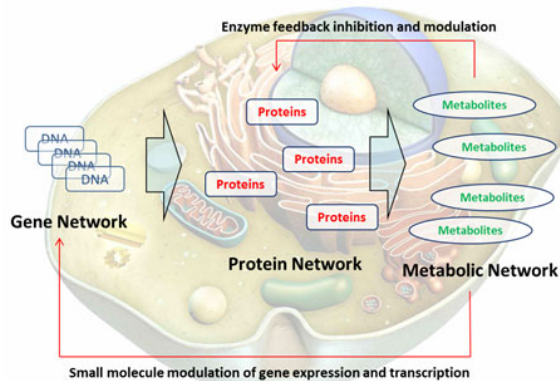
A number of **terms have evolved** that characterise these different elements and approaches within the metabolomics framework.

Metabolome: This refers to the entire set of metabolites found in a cell, tissue or organism.

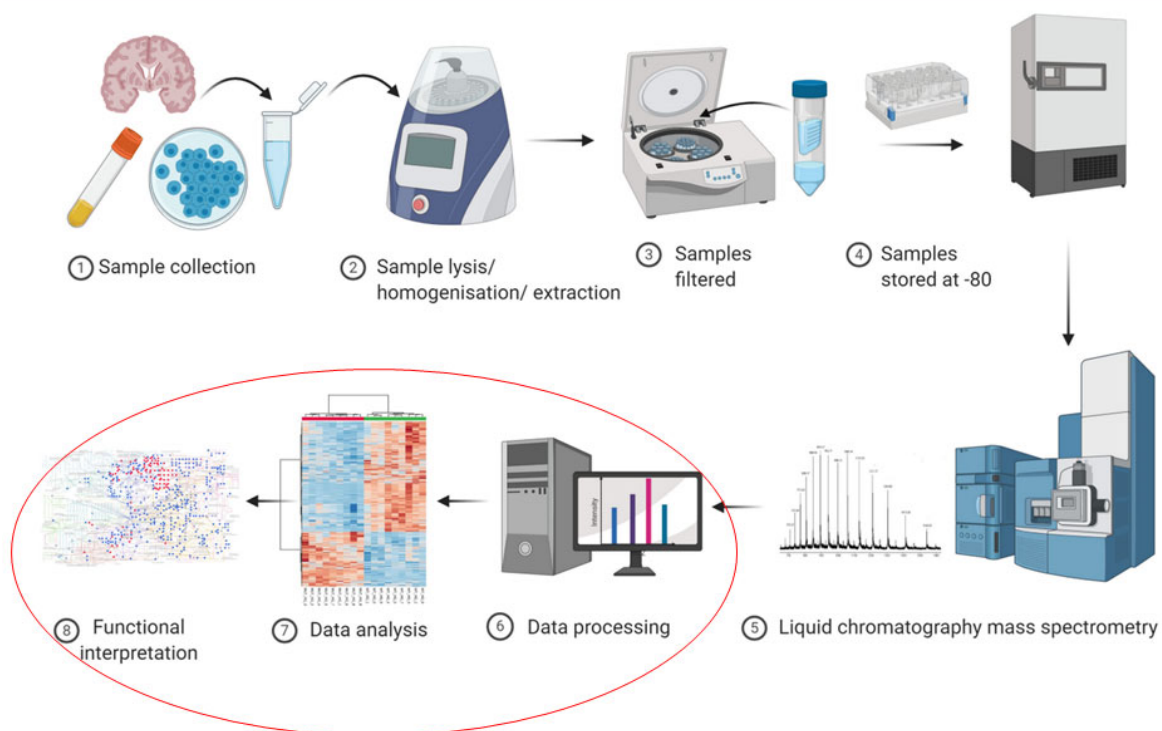
Metabolic profiling: The analysis of a selected number of pre-defined metabolites in a biochemical system.

Metabolic fingerprinting: The metabolites and their concentrations as a snapshot in time representing the state of a biological system or organism.

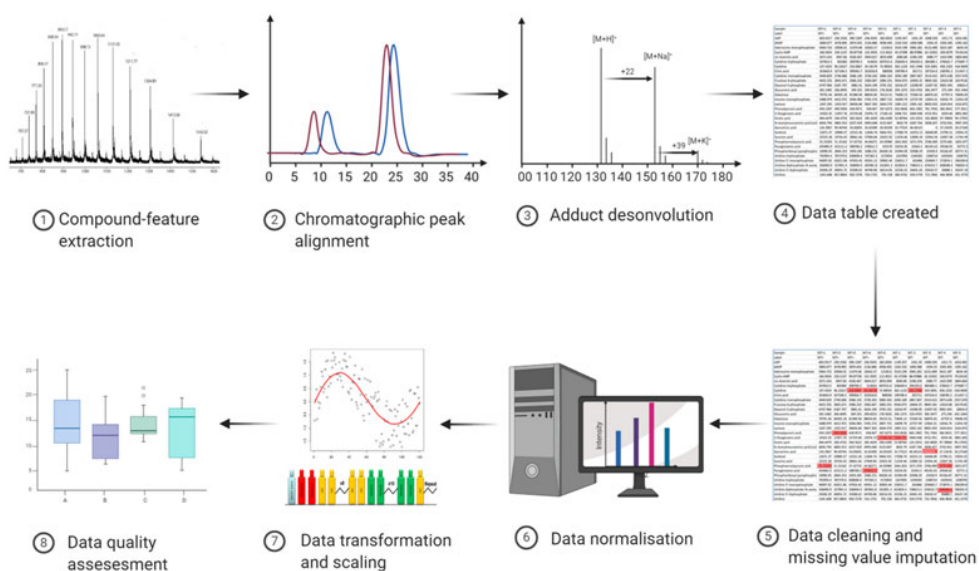
Metabolic footprinting: This refers to measuring what cells or a biological system excretes under controlled conditions, also known as the exometabolome.



OVERVIEW OF THE METABOLOMICS WORKFLOW



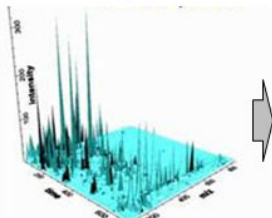
OVERVIEW OF DATA PROCESSING



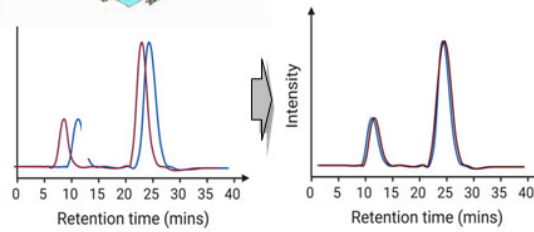
The **purpose of data processing** is to shape and format platform-specific raw data into a data table or matrix appropriate for statistical analysis. The process is split into two parts.

1. DATA (PRE) PROCESSING

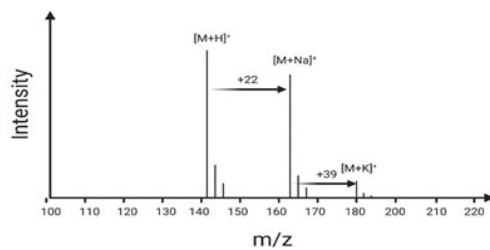
1. Spectral peak picking



2. Chromatographic peak alignment



3. Adduct deconvolution



Sample	WT-1	WT-2	WT-3	WT-4	WT-5	WT-2	WT-3	WT-4	WT-5	
Label	WT-1	WT-2	WT-3	WT-4	WT-5	WT-2	WT-3	WT-4	WT-5	
UDP	403.9237	236.1556	390.1587	236.9245	382.0824	149.357	1341.29	3468.539	1811.73	1853.882
ADP	308.0277	347.895	2874.655	2136.886	4938.45	1242.53	1405.588	1504.23	3305.45	1395.362
Adenosine monophosphate	9364.723	1306.05	11374.09	10263.37	11320.6	1929.339	1965.381	4315.699	5613.387	4634.34
Cyclic AMP	246.9204	150.1137	98.87728	152.8935	113.8212	43.7298	88.97864	63.16202	120.6709	79.02124
iso-Acetic acid	3571.424	3947.85	4330.467	3944.037	3876.699	3000.8	2396.239	2088.77	1423.999	1804.664
Cytidine triphosphate	3383.613	3020.8	30970.5	31664.697616	25668.4	294316.4	26688.4	294316.4	38488.4	27642.7
Cytidine	147.4224	96.21627	220.0867	34.18179	72.90264	363.1333	425.1948	529.3883	408.1334	434.9649
Citic acid	35866.8	327206.5	399663.7	352028.8	380258	397828.4	201711	372724.6	238970.2	211403.1
Cordosine monophosphate	3448.829	2786.686	3588.325	3728.263	3084.324	2030.389	2867.667	2514.613	3874.628	2917.426
Glyceral 3-phosphate	4422.295	2845.671	3088.332	2583.067	2083.231	1544.073	1004.25	9840.31	1203.08	15379.82
Glycerol 3-phosphate	6707.965	6227.37	3851.42	3634.199	3795.231	3533.87	3208.99	1207.92	3021.95	3202.84
Gluconic acid	365.1482	266.8495	349.325	209.4253	178.2626	294.1373	300.4709	306.3077	275.294	432.1844
Galactose	79791.46	8625.38	81389.84	8824.85	76333.51	78000.13	72389.56	66876.45	67707.4	26066.29
Inosine monophosphate	5488.979	4412.971	5064.963	576.274	2807.73	2609.79	3277.99	32824.21	2104.16	11013.58
Lactose	1267.291	135.937	2464.64	9667.30	2444.579	1005.212	2205.162	8902.033	1624.814	1833.875
Phenylpyruvic acid	429.1327	499.9384	418.9571	558.067	347.6273	652.0486	465.362	561.7662	571.2021	567.2021
2-Oxoglutaric acid	14322.35	1197.76	2173.68	1976.79	1738.42	1808.71	1004.938	4702.81	4038.46	3891.002
Citic acid	406.4479	240.4733	34.2422	185.4029	246.6399	1230.764	125.5533	502.8029	3788.8	98.17926
is-Acetylthioacetic acid (iso)	4056.794	4863.623	6227.609	4993.046	4123.847	3812.79	4387.764	3918.627	3732.831	3997.293
Quinolinic acid	124.3967	99.40784	54.65895	35.83389	24.02339	33.77623	40.40223	0	27.2476	20.27269
Sorbitol	12021.57	10989.27	12515.36	11604.74	11644.61	17908.79	16274.13	16648.89	12964.51	13324.85
Succinic acid	2725.38	1973.43	2864.63	27960.84	22053.92	11259.46	13000.18	12954.34	12007.38	11742.49
Phosphoenolpyruvic acid	25.22029	15.81647	57.67712	44.80771	24.90964	264.883	3275.814	2766.899	2770.881	1825.077
Pyruglutamic acid	454665.4	451111.2	588768.2	378012.2	305510	3154.92	3034.92	42648.63	42648.63	33772.1
Phosphoenolpyruvate	30096.91	2844.251	2493.285	3388.211	8420.16	13344.09	32098.29	24264.9	43346.87	28772.51
Uridine triphosphate	702991.4	767379.4	838688.4	747365.5	572884	1262931	1240283	1382714	1420561	1282976
Uridine 5'-monophosphate	94937.02	65021.86	47933.45	43931.12	30902.44	23811.7	263486	274981.7	236896	279481.1
Uridine diphosphate (isoant)	245691.5	327021.4	358426.9	287818.8	242018.5	616284.5	708615.5	676124.7	689894.4	746025.4
Uridine 5'-diphosphate	29206.29	40056.72	33308.62	44700.86	56154.05	22226.25	24401.43	33018.17	26888.1	33837.38
Uridine	1341.608	807.8054	900.7278	724.1795	792.158	687.472	516.5738	725.7066	466.984	431.3794
17.71_121.020(m/z)	394.2209	398.206	1461.398	800.82	62.268	46.7898	50.81884	43.47028	78.38163	41.02022
17.13_160.040(m/z)	15427.79	902.41	6947.999	5962.297	2781.14	54348.54	43300.55	43498.63	35789.65	41743.48
14.87_184.040(m/z)	285.103	2601.88	2834.242	2628.215	2123.711	17775.81	30789.83	25444.51	21649.9	14862.8
11.41_175.060(m/z)	284.1571	396.4337	45.7624	60.89247	42.62255	81.6208	45.82114	26.51894	0	20.29644
21.18_165.060(m/z)	512.905	520.608	670.221	413.9882	186254.4	246739.9	276013.3	233022	262255	367061
41.15_166.040(m/z)	249.101	214.8	244.48	2222.761	1794.462	2605.268	12046.42	26027.86	29027.86	3113.861
9.11_229.021(m/z)	874.129	488.438	6409.328	4737.07	4032.365	14894.18	14488.95	14980.04	12879.45	15249.25
64.26_131.040(m/z)	8319.123	7589.966	8622.475	6023.463	5189.544	10873.61	11068.49	12466.78	10755.41	11837.12
10.19_209.020(m/z)	532.2064	511.4664	446.6409	405.2155	483.983	560.1833	422.7872	544.7377	561.966	472.9183
40.970_161.020(m/z)	429070.36	310461.1	47392.07	20873.63	44028.9	46465.22	51037.87	51460.79	59565.88	48807
20.56_131.040(m/z)	2257.623	2975.737	2292	2260.803	2292.362	1523.523	1861.609	1915.184	1582.186	1189.262
24.34_204.040(m/z)	107.4429	102.3042	308.7847	238.5233	283.0005	1113.751	1298.528	1172.777	980.836	1003.674
15.12_179.040(m/z)	421.8997	432.6325	441.1304	724.092	220.6099	138.445	707.8614	614.8182	548.4	455.9224
18.47_178.020(m/z)	1313.348	1361.276	1181.085	1000.072	702.7288	792.4916	638.3973	781.1517	844.9632	544.4034
16.11_131.040(m/z)	6132.03	7902.61	9504.49	7817.9	6864.91	5662.46	6127.19	4425.16	6120.61	2004.46
11.15_129.020(m/z)	657.883	700.068	544.284	577.6549	560.792	594.066	646.1211	584.9034	574.9642	471.7678
30.06_145.020(m/z)	4990.123	3891.821	4501.808	5006.008	4120.55	4054.941	3807.386	4654.464	3782.17	2899.838
21.06_201.030(m/z)	980.245	1453.648	2348.475	3625.264	874.2327	655.9944	618.1229	789.6163	744.623	594.9796
14.58_229.021(m/z)	465.4284	734.471	841.9008	913.79	548.611	2770.439	4708.413	3074.65	2778.071	5187.238
21.27_179.020(m/z)	22376.76	7942.767	4784.822	7168.162	4086.154	10563.47	11212.18	10977.62	10244.81	12046.81
23.18_426.022(m/z)	8677484	3869889	5808311	3018768	10025772	1930461	2202228	1689977	2193211	2193475

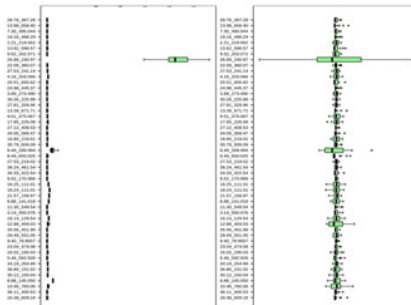
Peak intensity or concentration table

Sample	WT-1	WT-2	WT-3	WT-4	WT-5	WT-2	WT-3	WT-4	WT-5	
Label	WT-1	WT-2	WT-3	WT-4	WT-5	WT-2	WT-3	WT-4	WT-5	
UDP	403.9237	236.1556	390.1587	236.9245	382.0824	149.357	1341.29	3468.539	1811.73	1853.882
ADP	308.0277	347.895	2874.655	2136.886	4938.45	1242.53	1405.588	1504.23	3305.45	1395.362
Adenosine monophosphate	9364.723	1306.05	11374.09	10263.37	11320.6	1929.339	1965.381	4315.699	5613.387	4634.34
Cyclic AMP	246.9204	150.1137	98.87728	152.8935	113.8212	43.7298	88.97864	63.16202	120.6709	79.02124
iso-Acetic acid	3571.424	3947.85	4330.467	3944.037	3876.699	3000.8	2396.239	2088.77	1423.999	1804.664
Cytidine triphosphate	3383.613	3020.8	30970.5	31664.697616	25668.4	294316.4	26688.4	294316.4	38488.4	27642.7
Cytidine	147.4224	96.21627	220.0867	34.18179	72.90264	363.1333	425.1948	529.3883	408.1334	434.9649
Citic acid	35866.8	327206.5	399663.7	352028.8	380258	397828.4	201711	372724.6	238970.2	211403.1
Cordosine monophosphate	3448.829	2786.686	3588.325	3728.263	3084.324	2030.389	2867.667	2514.613	3874.628	2917.426
Glyceral 3-phosphate	4422.295	2845.671	3088.332	2583.067	2083.231	1544.073	1004.25	9840.31	1203.08	15379.82
Glycerol 3-phosphate	6707.965	6227.37	3851.42	3634.199	3795.231	3533.87	3208.99	1207.92	3021.95	3202.84
Gluconic acid	365.1482	266.8495	349.325	209.4253	178.2626	294.1373	300.4709	306.3077	275.294	432.1844
Galactose	79791.46	8625.38	81389.84	8824.85	76333.51	78000.13	72389.56	66876.45	67707.4	26066.29
Inosine monophosphate	5488.979	4412.971	5064.963	576.274	2807.73	2609.79	3277.99	32824.21	2104.16	11013.58
Lactose	1267.291	135.937	2464.64	9667.30	2444.579	1005.212	2205.162	8902.033	1624.814	1833.875
Phenylpyruvic acid	429.1327	499.9384	418.9571	558.067	347.6273	652.0486	465.362	561.7662	571.2021	567.2021
2-Oxoglutaric acid	14322.35	1197.76	2173.68	1976.79	1738.42	1808.71	1004.938	4702.81	4038.46	3891.002
Citic acid	406.4479	240.4733	34.2422	185.4029	246.6399	1230.764	125.5533	502.8029	3788.8	98.17926
is-Acetylthioacetic acid (iso)	4056.794	4863.623	6227.609	4993.046	4123.847	3812.79	4387.764	3918.627	3732.831	3997.293
Quinolinic acid	124.3967	99.40784	54.65895	35.83389	24.02339	33.77623	40.40223	0	27.2476	20.27269
Sorbitol	12021.57	10989.27	12515.36	11604.74	11644.61	17908.79	16274.13	16648.89	12964.51	13324.85
Succinic acid	2725.38	1973.43	2864.63	27960.84	22053.92	11259.46				

2. DATA PROCESSING: TRANSFORMATION AND SCALING

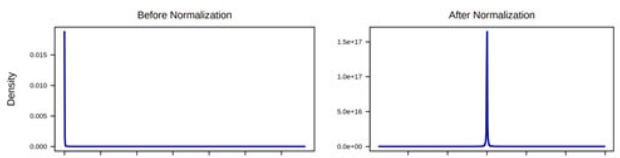
Data table

Sample	WT-1	WT-2	WT-3	WT-4	WT-5	WT-1	WT-2	WT-3	WT-4	WT-5
Label	WT+	WT+	WT+	WT+	WT+	WT-	WT-	WT-	WT-	WT-
ADP	603.9217	230.1506	390.1587	236.5095	382.8924	1149.937	1341.29	3408.559	5413.71	1603.481
Adenosine monophosphate	594.723	1506.051	11374.06	10283.57	11530.6	5129.299	5965.281	6131.899	5615.387	8634.34
Cyclic ADP	585.0564	501.1517	95.89728	152.3915	113.8522	18.37298	18.97986	62.3182	269.0279	79.02124
α-Amino acid	5372.424	1947.85	6208.467	3944.017	3876.659	2000.89	2198.239	2089.77	1423.999	1804.664
Cytidine triphosphate	337813.5	30282	30790.5	134054	697925.6	256691.4	294293.6	294391.1	278422.7	275687.7
Cytidine	147.4254	95.22617	380088	183328	1930024	262.1219	3800286	29.3601	456.1253	425.9699
Citric acid	303663.9	527226.5	399363.7	310251.8	880058	399368.4	201711	397224.4	238795.2	211457.1
Cytidine monophosphate	1440.829	2740.695	5086.205	3728.263	3084.234	2030.389	2667.427	2144.615	2875.428	2017.676
Fructose 6-phosphate	4422.255	2864.471	1706.212	2281.047	2083.211	9544.073	34004.02	9840.182	12010.68	1079.842
Glyceral 3-phosphate	6247.966	6187.707	5881.41	5634.299	3795.211	16326.17	13298.99	13297.10	9005.945	10820.4
Glucuronic acid	361.842	356.8695	389.225	209.8253	176.2626	294.1273	125.4703	106.3677	175.294	402.1464
Galactose	79795.46	84405.18	81389.34	88024.85	76153.51	73600.13	72389.56	66476.65	67507.4	70066.03
Inosine monophosphate	5488.979	4412.92	5046.983	5765.274	2607.715	5609.79	62737.99	13824.21	10556.74	11053.39
Lactone	1292.795	1313.627	24654.44	9402.93	2644.729	405.212	2905.142	8602.019	1044.814	1814.475
Phenylethylamine	409.1267	370.76	418.9731	558.027	347.4273	612.0636	665.1902	701.7406	542.8421	577.2011
2-Oxoglutaric acid	14322.35	11957.76	12729.68	10795.71	7899.9992	5644.98	4722.161	6559.46	3801.082	3077.3
Oxetic acid	404.479	240.4733	364.5422	185.4429	246.4189	52.00764	125.5533	302.8828	87.78904	99.17953
N-Acetylmuramic acid	4056.794	4865.613	6227.609	4993.048	4123.847	382.79	4837.764	3938.427	3732.881	3997.293
Quinic acid	124.967	94.40284	64.94065	76.81039	34.02109	13.79215	48.80219	37.29426	29.37289	37.0788
Sorbitol	12471.37	10989.27	12555.36	11604.74	9464.511	17908.74	34253.13	36648.89	15796.51	11824.25
Isocitric acid	22258.38	19726.43	20943.66	27989.84	22025.10	11259.46	12000.18	12594.18	12027.48	11742.49
Phosphoenolpyruvic acid	470485.9	413111.2	508768.2	395570	31254.10	13341.5	40145.43	29748.05	10772.3	14157.07
Pyruvic acid	3006.95	3844.253	3439.205	308.221	84020.16	13944.09	37008.29	24248.9	41656.47	39772.51
Uridine triphosphate	79293.4	787379.6	638688.4	747935.5	132954	1267099	1249283	1308714	140543	120879
Uridine 5'-monophosphate	94097.02	40221.86	47933.45	49315.12	20002.44	238913.7	263446	216889.7	273781.2	246099.8
Uridine diphosphate N-acid	204869.9	877901.4	730640.9	879461.8	241891.3	432024.5	798161.5	4761032	3800000	240201.4
Uridine 5'-diphosphate	27036.29	40055.72	33308.62	44700.86	56164.05	22326.25	24021.43	33028.17	20888.1	13037.38
Uridine	1341.628	857.8854	950.9729	724.1255	792.158	4874792	3583578	725.7855	448.9834	423.3719



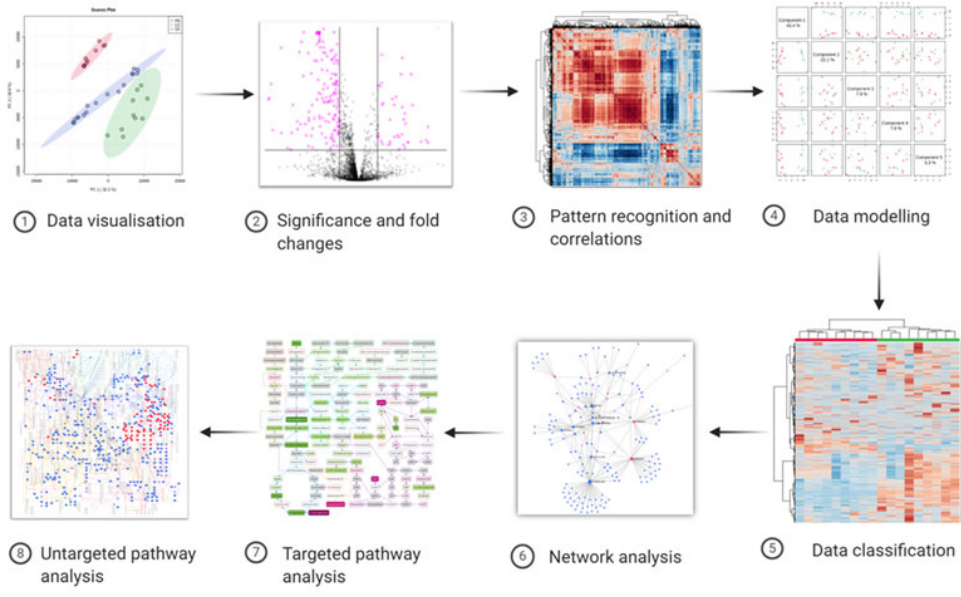
Data scaling: brings metabolite abundances onto a similar scale

Data transformation: ensures data is normally distributed



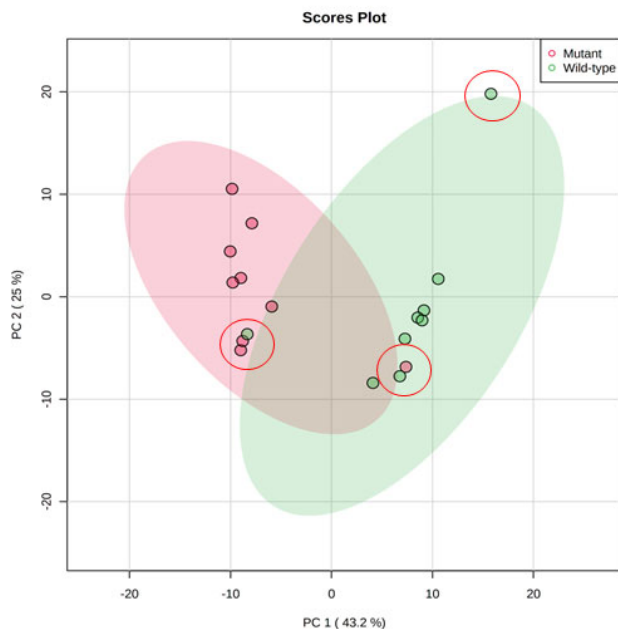
Class	Method	Formula	Unit	Goal	Advantages	Disadvantages	
I	Centering	$\bar{x}_{ij} = x_{ij} - \bar{x}_i$	0	Focus on the differences and not the similarities in the data	Remove the offset from the data	When data is heteroscedastic, the effect of this pretreatment method is not always sufficient	
	II	Autoscaling	$\bar{x}_{ij} = \frac{x_{ij} - \bar{x}_i}{s_i}$	(-)	Compare metabolites based on correlations	All metabolites become equally important	Inflation of the measurement errors
		Range scaling	$\bar{x}_{ij} = \frac{x_{ij} - \bar{x}_i}{(x_{i,max} - x_{i,min})}$	(-)	Compare metabolites relative to the biological response range	All metabolites become equally important. Scaling is related to biology	Inflation of the measurement errors and sensitive to outliers
		Pareto scaling	$\bar{x}_{ij} = \frac{x_{ij} - \bar{x}_i}{\sqrt{s_i}}$	0	Reduce the relative importance of large values, but keep data structure partially intact	Stays closer to the original measurement than autoscaling	Sensitive to large fold changes
		Vast scaling	$\bar{x}_{ij} = \left(\frac{x_{ij} - \bar{x}_i}{s_i} \right) \frac{\bar{x}_i}{s_i}$	(-)	Focus on the metabolites that show small fluctuations	Aims for robustness, can use prior group knowledge	Not suited for large induced variation without group structure
III	Level scaling	$\bar{x}_{ij} = \frac{x_{ij} - \bar{x}_i}{\bar{x}_i}$	(-)	Focus on relative response	Suited for identification of e.g. biomarkers	Inflation of the measurement errors	
	Log transformation	$\bar{x}_{ij} = 10 \log(x_{ij})$ $\bar{x}_{ij} = x_{ij} - \bar{x}_i$	Log 0	Correct for heteroscedasticity, pseudo scaling. Make multiplicative models additive	Reduce heteroscedasticity, multiplicative effects become additive	Difficulties with values with large relative standard deviation and zeros	
	Power transformation	$\bar{x}_{ij} = \sqrt{x_{ij}}$ $\bar{x}_{ij} = x_{ij} - \bar{x}_i$	0	Correct for heteroscedasticity, pseudo scaling	Reduce heteroscedasticity, no problems with small values	Choice for square root is arbitrary	

DATA ANALYSIS AND INTERPRETATION OVERVIEW



The **purpose of data analysis** in metabolomics is to identify patterns and structures in the data which reveal biologically meaningful information.

DATA OVERVIEW: PRINCIPLE COMPONENTS ANALYSIS (PCA)



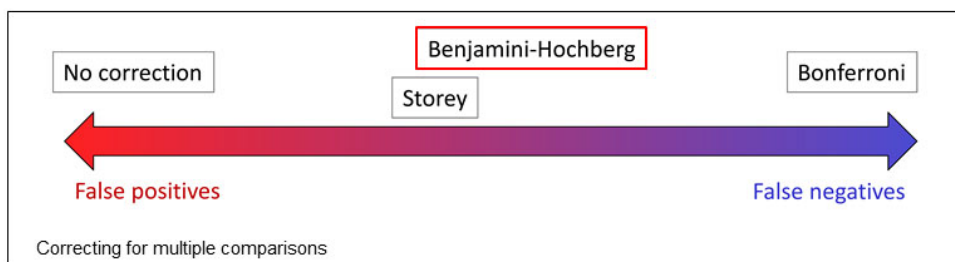
PCA in metabolomics

- Commonly applied as an overview of the dataset.
- Converts metabolite abundances into coordinates in multi-dimensional space with one principle component for each variable.
- Principle components maximise abundance variation in the data which generally decreases as the principle components increase.
- The first few principle components therefore usually show the maximal variation in the dataset.
- **Used to identify outlier samples or variables.**

IDENTIFYING THE SIGNIFICANCE OF METABOLITE CHANGES USING UNIVARIATE STATISTICS

$$\text{Fold change (FC)} = \frac{\text{Final concentration}}{\text{Initial concentration}}$$

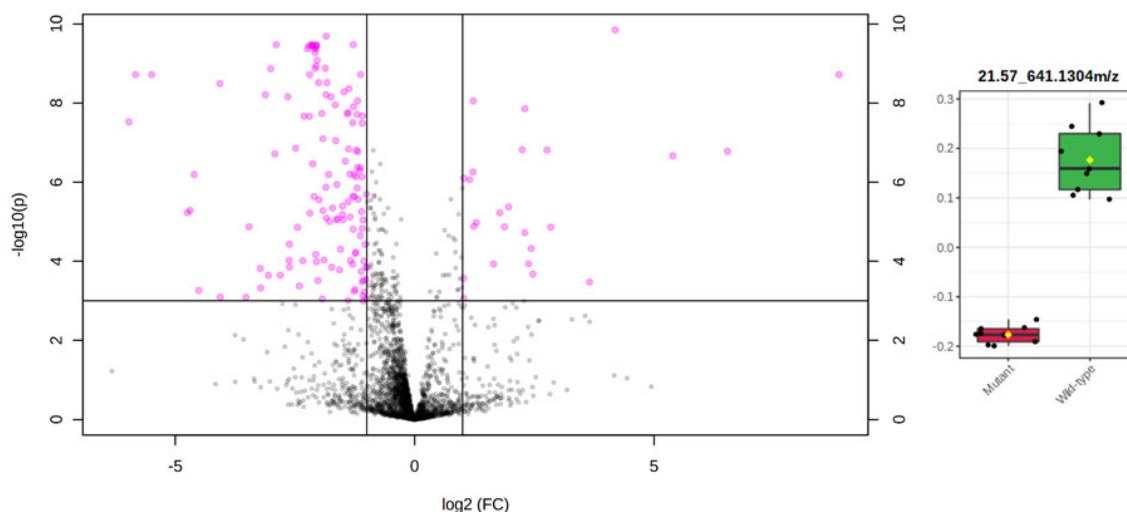
Type	Paired	Unpaired	Assumptions
Parametric	Student's t-test	Welch's t-test	The experimental population represents the 'true' population and is normally distributed
Non-parametric	Wilcoxin-signed rank	Mann-Whitney-U test	Does not assume the experimental population is normally distributed
ANOVA	Repeated measures ANOVA	One-way ANOVA	Analysis of variance (ANOVA) generalizes the t-test when the data belong to more than two groups. Assumes a normal distribution



"Have compound-features been modified by induced experimental differences and if so are these significant?"

VOLCANO PLOT

“How to select metabolites which are altered most significantly between experimental groups?”

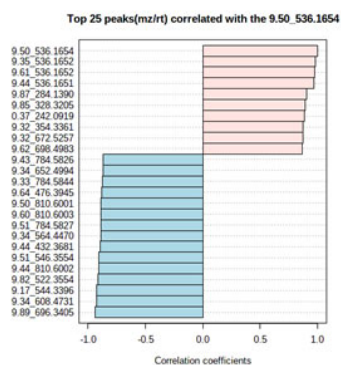


Some caution should be taken as **not all metabolites are regulated to the same extent**. Some pathways are more tightly regulated than others and hence a 1.5 fold change in one may be more important than a 2-fold change in another.

PATTERN RECOGNITION

Statistical approaches that calculate **correlation coefficients** can be used to identify the strength of linear relationships between variables. These commonly use **Pearson's r** or **Spearman's rank coefficients** to identify which variables correlated in the way they can in response to a change in biological condition. Correlation analysis can be used to answer simple questions such as **which metabolites also increase (or decrease) significantly when citrate levels go up in hypoxic environments?** The same approaches can also be used to identify more complicated relationships within the data, for examples which metabolites shows a similar pattern over a time-course.

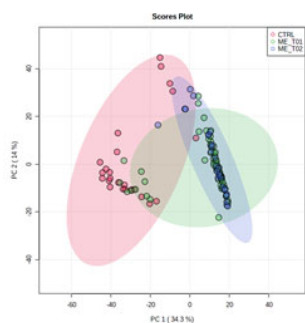
Although correlations can provide powerful information they are usually **sensitive to outliers, unequal variances and non-normality**. It should always be remembered that **correlation is not causation** and this approach does not provide any information about why the metabolite changes are correlated.



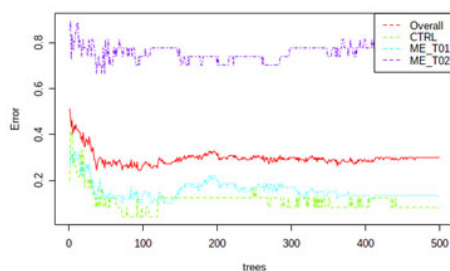
SUPERVISED MULTIVARIATE MODELLING OF METABOLOMICS DATA

Although univariate statistical analysis can help **identify metabolites that are significantly different** in abundance between experimental groups, univariate statistics ignores relationships between metabolites within datasets, they treat each compound-feature or metabolite as an independent variable.

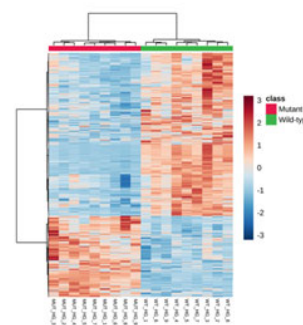
In reality we know this is not the case, that **many metabolites are inter-dependent** in a biological context and multivariate statistical approaches enables such relationships to be modelled as part of the data analysis: The benefit is that i) The analysis provides greater biological insight. ii) The reliability of the models can be statistically evaluated and ii) the models can be used prospectively for future studies.



Vector based approaches

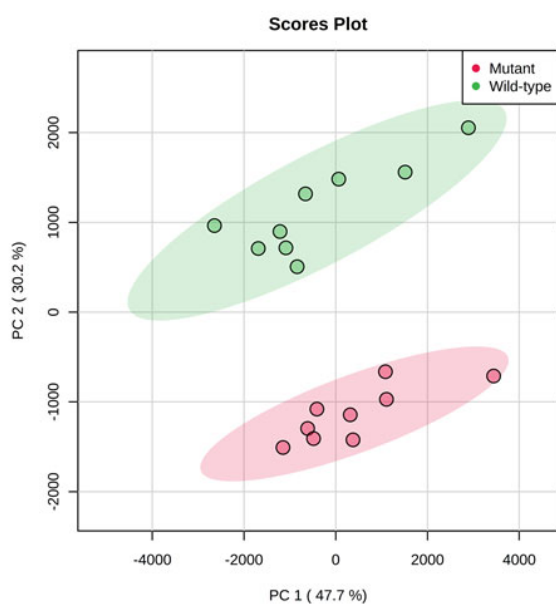


Random Forest classification



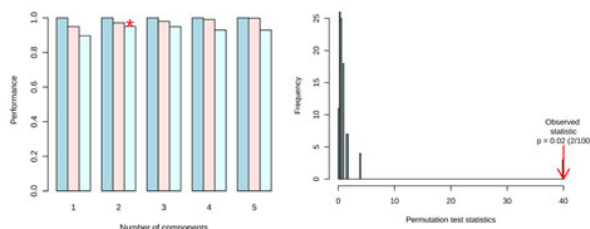
Hierarchical clustering

PARTIAL LEAST SQUARES DISCRIMINANT ANALYSIS (PLS-DA)



PLS-DA cross validation details:

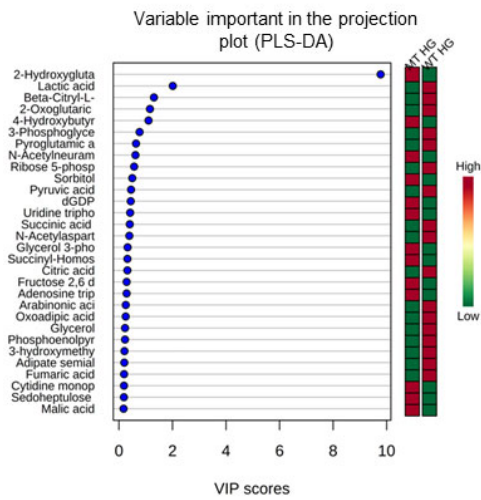
Measure	1 comps	2 comps	3 comps	4 comps	5 comps
Accuracy	1.0	1.0	1.0	1.0	1.0
R2	0.9494	0.97163	0.98006	0.99116	0.99821
Q2	0.89664	0.95087	0.94898	0.9297	0.92962



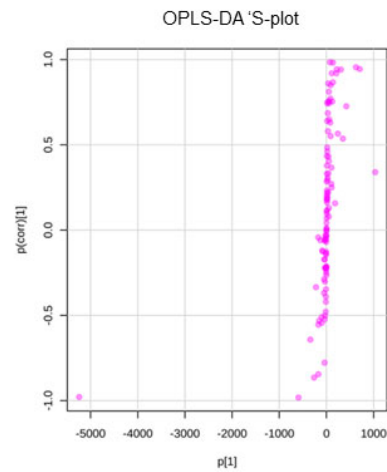
PLS-DA: A vector-based approach (like PCA) but rather than capturing only variance in the variables it also capture the covariance in the sample grouping.

Note: supervised multivariate techniques such as PLS-DA try to discriminate between experimental increasing the probability of doing so by chance. This can lead to **overfitting of data** and its important to **validate the models**.

PARTIAL LEAST SQUARES DISCRIMINANT ANALYSIS (PLS-DA)



When PLS-DA or OPLS-DA models have been constructed and validated they can be used to identify important variables.

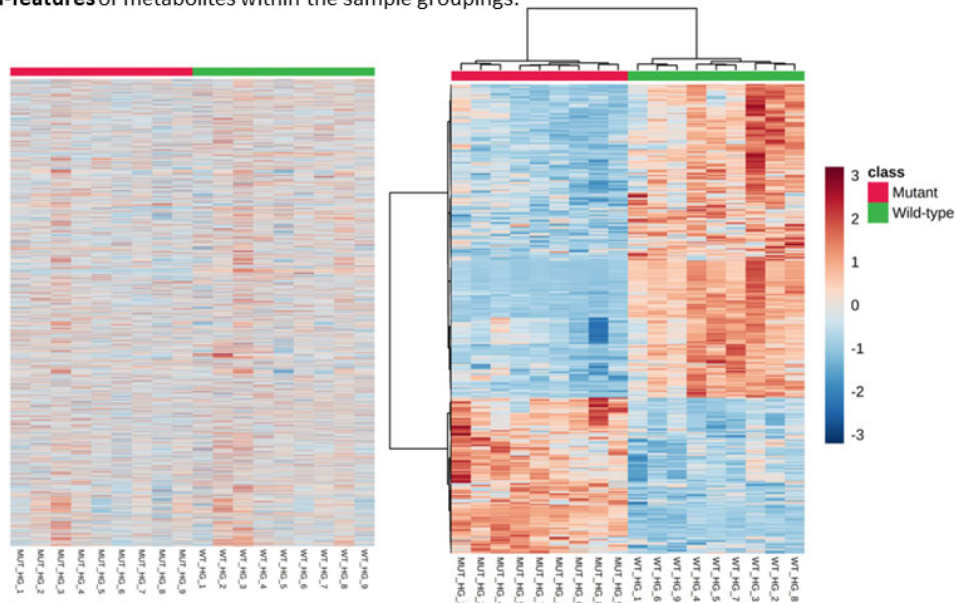


A variable importance in the projection plot (VIP plot) scores each variable according to the weighted sum of the squares of the PLS loading determined by the amount of explained variance between the experimental groups.

An S-plot combines the covariance and correlations. The Y-axis is the metabolite's reliability score in the model and on the X-axis is the direction and magnitude of its deviation from the control sample.

HIERARCHICAL CLUSTERING

Unsupervised multivariate statistical approach used to **organise complex datasets**, often visualised using heatmaps. **Dendrograms are often associated with these heat-maps to illustrate the similarity between the different compound-features** or metabolites within the sample groupings.



Identifying relationships between metabolites

PUTTING DATA PROCESSING AND ANALYSIS INTO PRACTICE

[MeaboAnalyst: https://www.metaboanalyst.ca/](https://www.metaboanalyst.ca/)

MetaboAnalyst is a free online data processing and analysis platform provided for the metabolomics community by the Xia Lab @ McGill university in Canada. It provides software tools for data processing, analysis, visualisation and functional interpretation of targeted and untargeted metabolomics data.