Metabolomics Summary

Definitions of metabolomics

Metabolites = the **small molecules** (<1000 Da) found within a biological system at specific time under specific condition

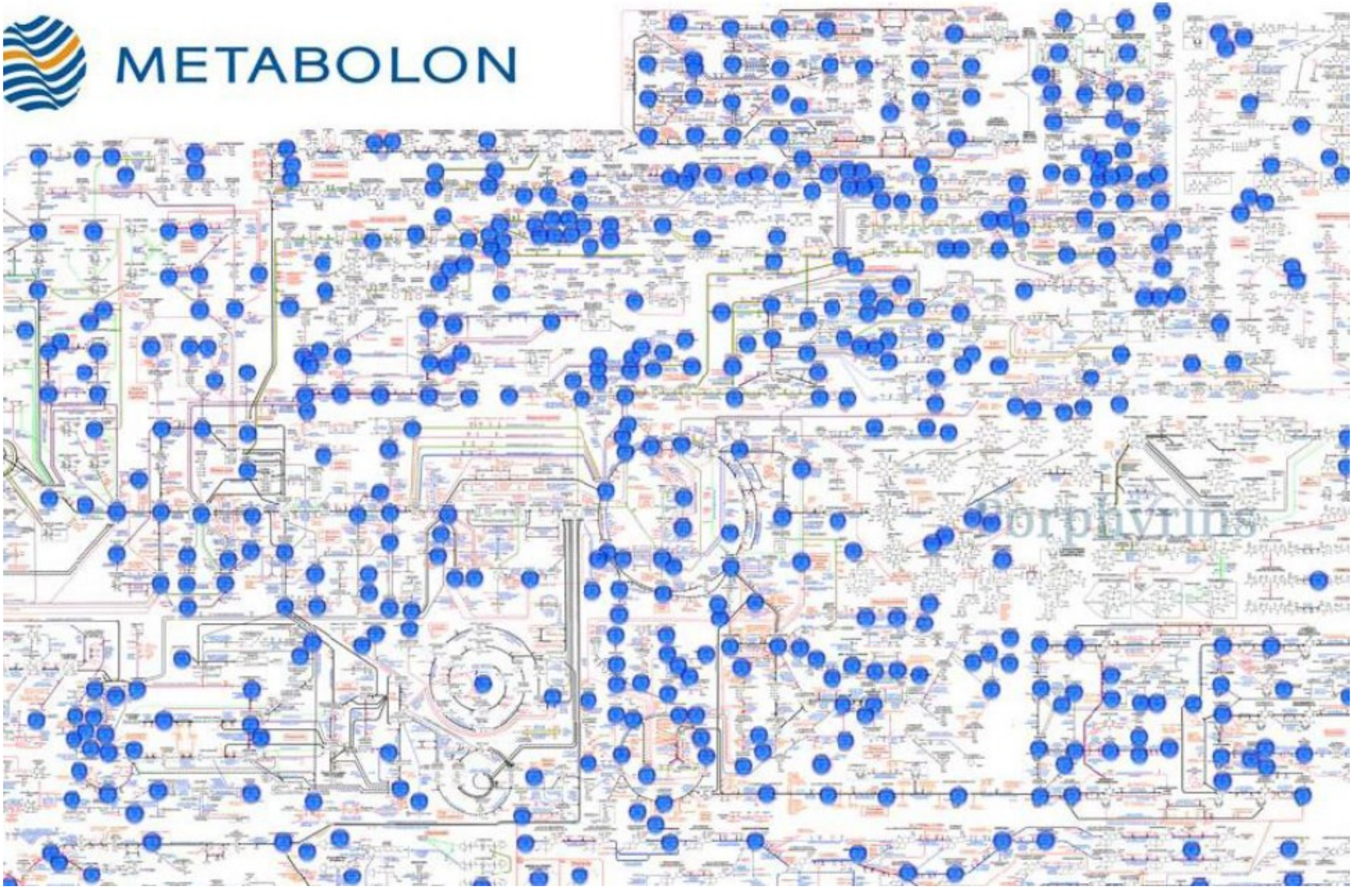
- true **endpoints** of most biological processes

Metabolome = complete set of all metabolites

- around ~2500 (+~3500 food +~1200 drugs)

Metabolomics = detection, quantitative measurement and **analysis of metabolomes**

How to measure metabolites?



Sample Collection

- **Blood**
- **Urine**
- **Further body fluids** (Cerebrospinal Fluid , Saliva , Sweat , Tears , Feces , Breath air)
- **Tissues** (liver, muscle, kidney, fat)
- **Cell cultures**
- **Plant extract**

temperature, addictive and lab difference could affect the measu

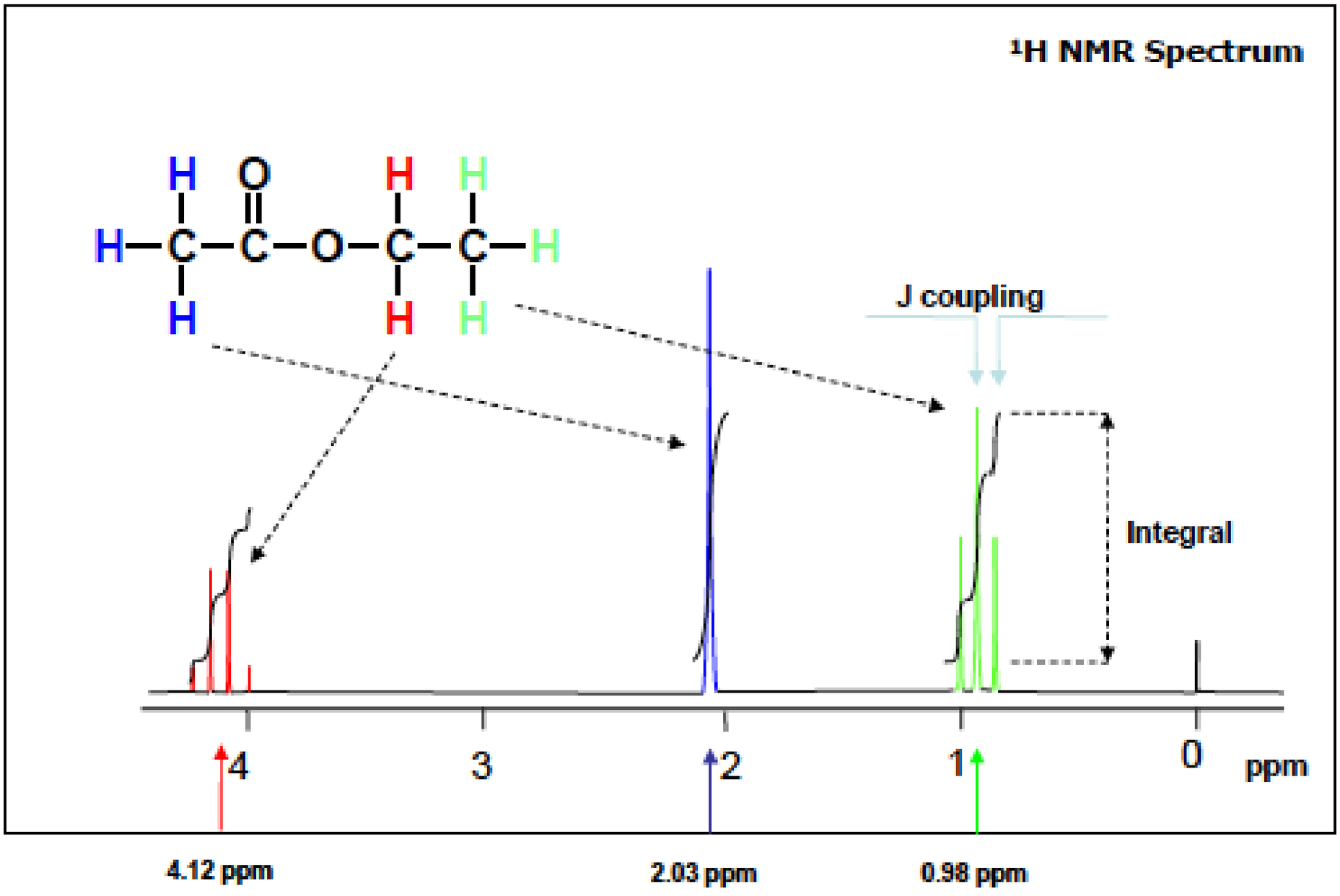
Metabolite detection and quantification

Targeted	Non-targeted
Preselected set of signals	All signals
Pros: Known identity; better quantification	Pros: New/Unknown metabolites
Cons: No new metabolites	Cons: Difficult identification; less reliable quantification
Routine	**Discovery**

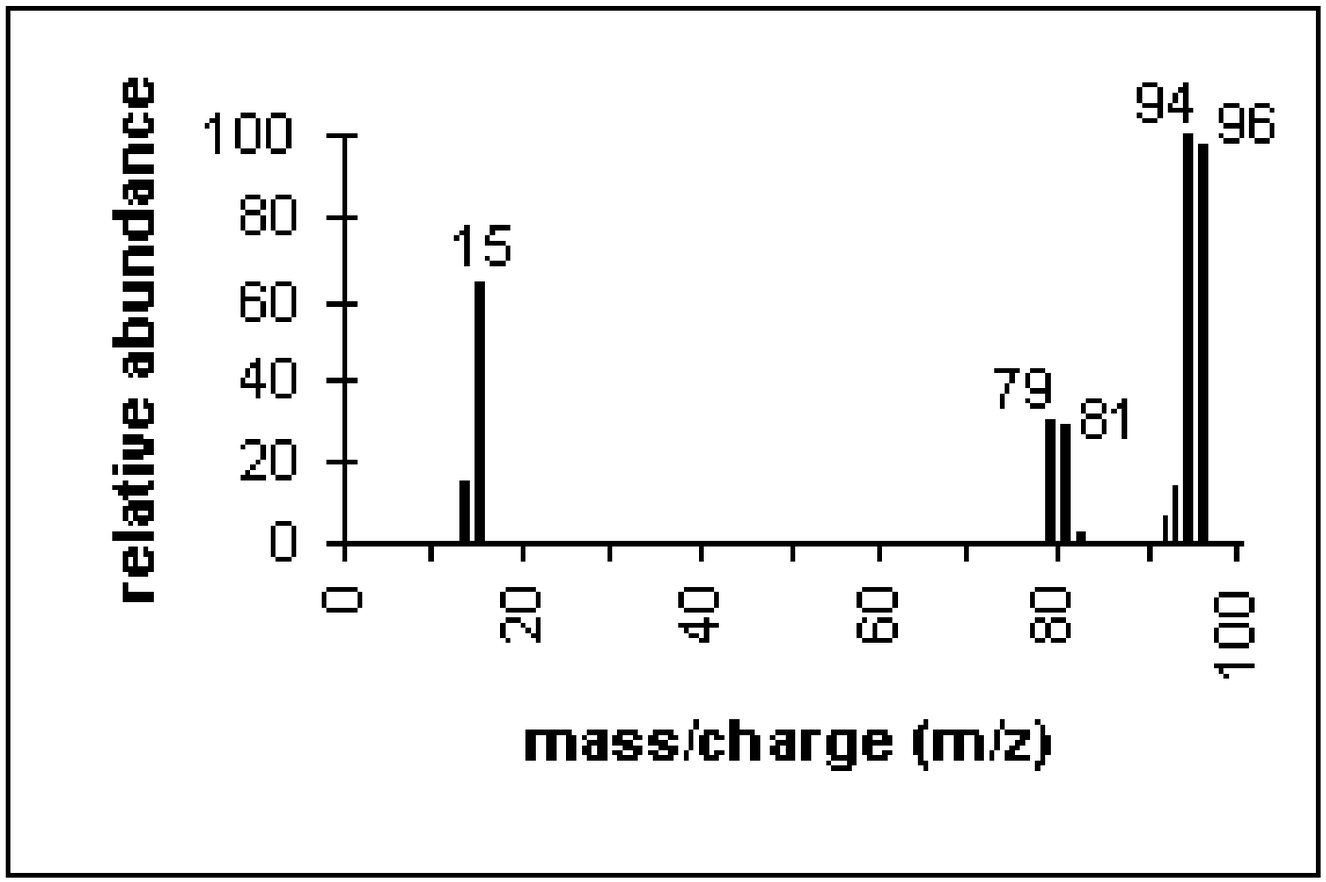
Technologies of high-throughput metabolomics

NUCLEAR MAGNETIC RESONANCE (NMR)	MASS SPECTROMETRY (MS)
energy changes during resonance of molecules	measure mass of the fragments of molecules
ppm (parts per million)	Pros: more sensitive
Pros: no pre-processing, more stable	Cons: less stable (noise)
Cons: less sensitive	

NUCLEAR MAGNETIC RESONANCE (NMR)



MASS SPECTROMETRY (MS)
Relative abundance of isotopes



$Br{-}CH_3$

m/z = 15

(79)BrCH3 m/z = 94

(81)BrCH3 m/z = 96

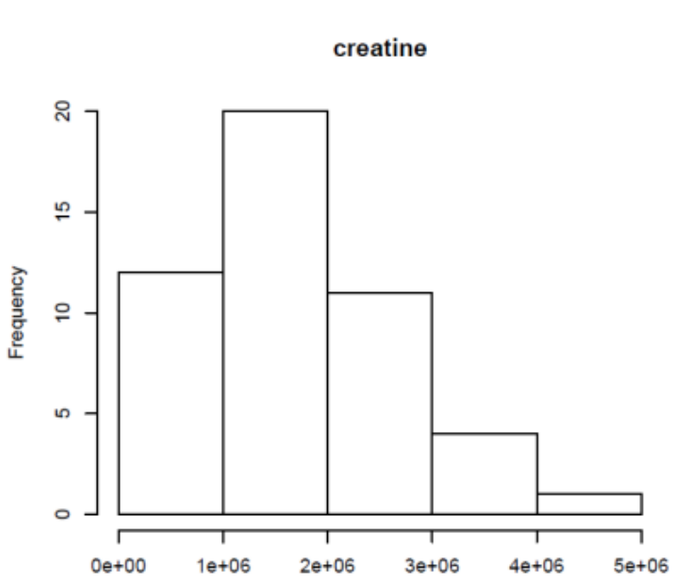
Data Analysis — Four steps

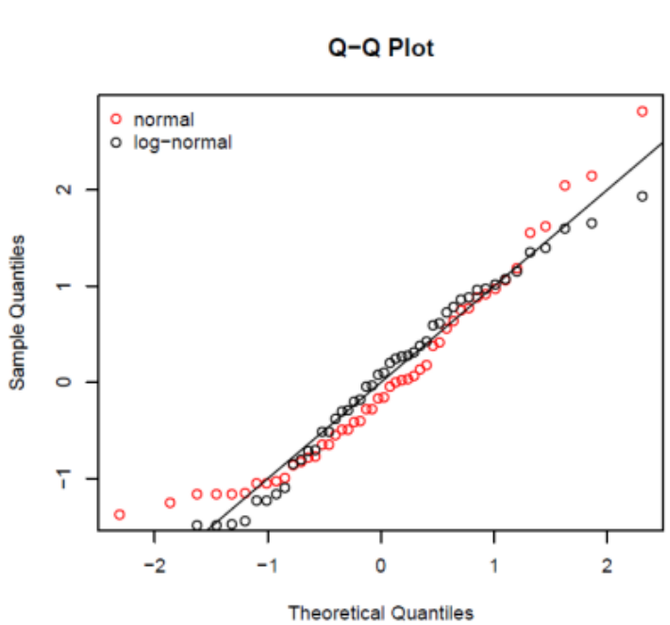
1/ **Raw data processing** — peak detection, peak alignment, peak integration, identification of metabolites, …

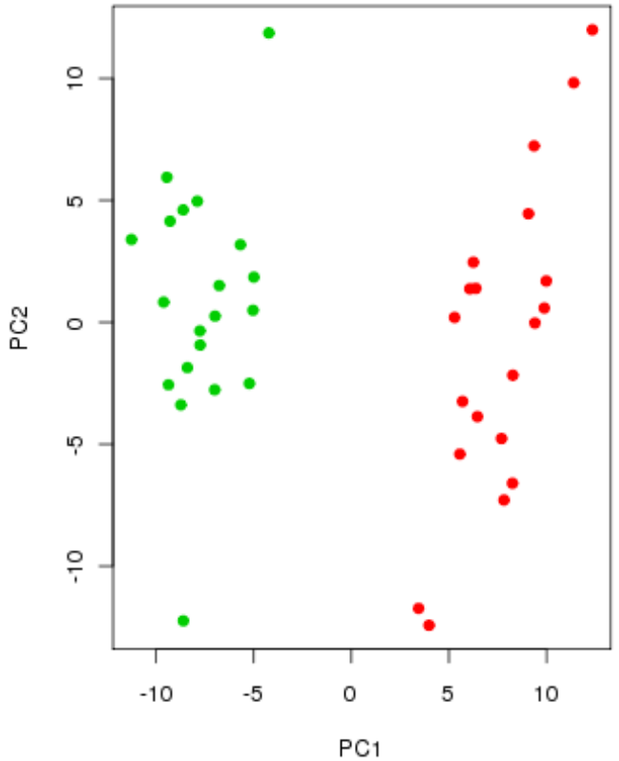
	A	B	C	D	E	F	G	H
1		alanine	asparagine	aspartate	beta-alanine	2-aminobutyrate	creatine	creatinine
2	KEGG_ID	C00041	C00152	C00049	C00099	C02356	C00300	C00791
3	G49	132916896	1108448	38198356	1660502	210544	2103232	167176
4	G50	133477084	914052	43682177	1347281	218637	1783242	109076
5	G51	136103631	775144	40335416	996984	119311	2792832	109632
6	G52	132434066	944874	34257895	1156098	263267	1148866	93976
7	G53	118492860	470526	30985593	1427715	206972	1474555	25578
8	G54	125060695	787470	30998311	1594232	136421	1481915	
9	G55	111198318	1140890	37525785	1912930	288790	2547325	104856
10	G56	106483460	726200	31669514	1660297	279919	1892293	
11	G57	115102642	1322471	33774414	1260397	177708	1686154	
12	G58	123560493	498869	30368312	991845	178753	2233373	57443
13	G59	124807228	1081946	32261693	1301586	110430	1722994	144087
14	G60	131011468	1266590	36541090	839685	137570	1234340	47371
15	G61	169107274	728200	37304593	778996	87187	957581	
16	G62	161757510	1453406	37932822	1432621	68507	2396696	98745
17	G63	166919106	1085864	31778624	1394881	72894	682267	109200
18	G64	119386346	1099144	37773027	1672703	203888	982965	
19	G65	144521192	695337	36696010	1613429	160347	1352265	73279

Data Analysis — Four steps

2/ **Primary data analysis (QC)**: outlier detection, normalization (batch effects, dilution), missing value handling/imputing…



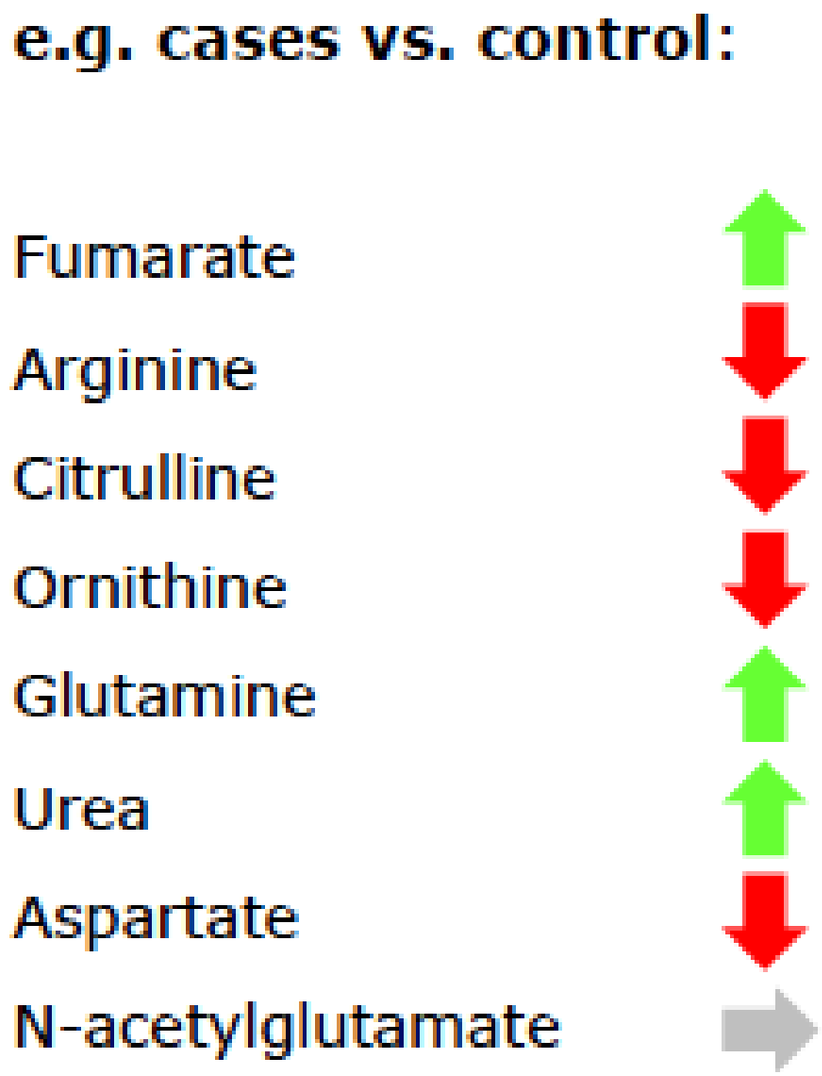




Data Analysis — Four steps

3/ **Statistical analysis**:

univariate/multivariate hypothesis tests, supervised/unsupervised machine learning (classification/clustering), …

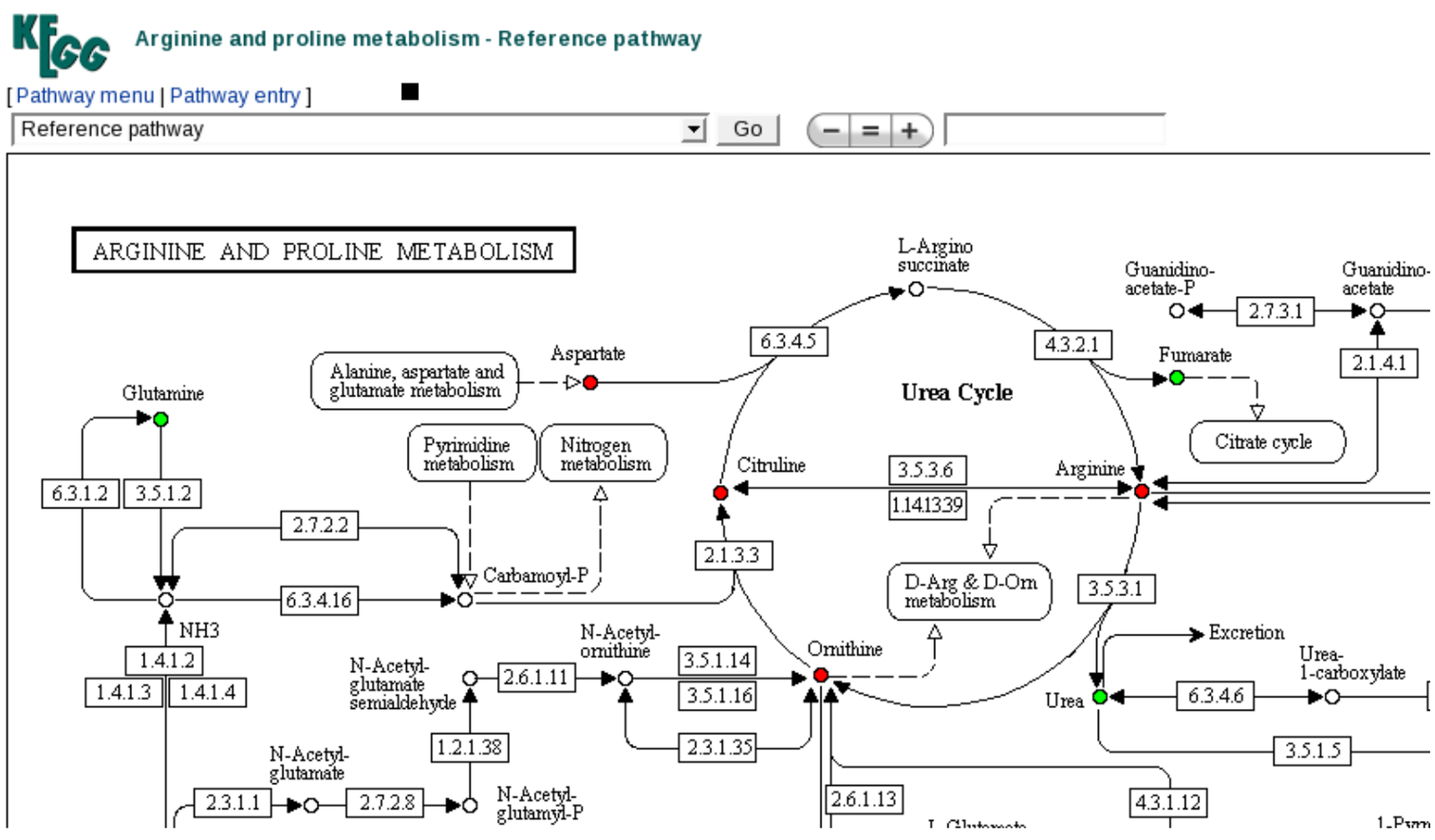


Data Analysis — Four steps

4/ **Bioinformatic analysis**: biological context, network analyses, data integration

(1) Mapping

- not all metabolites are measured
- Measured and map metabolites do not match exactly
- No mapping of unknown metabolites



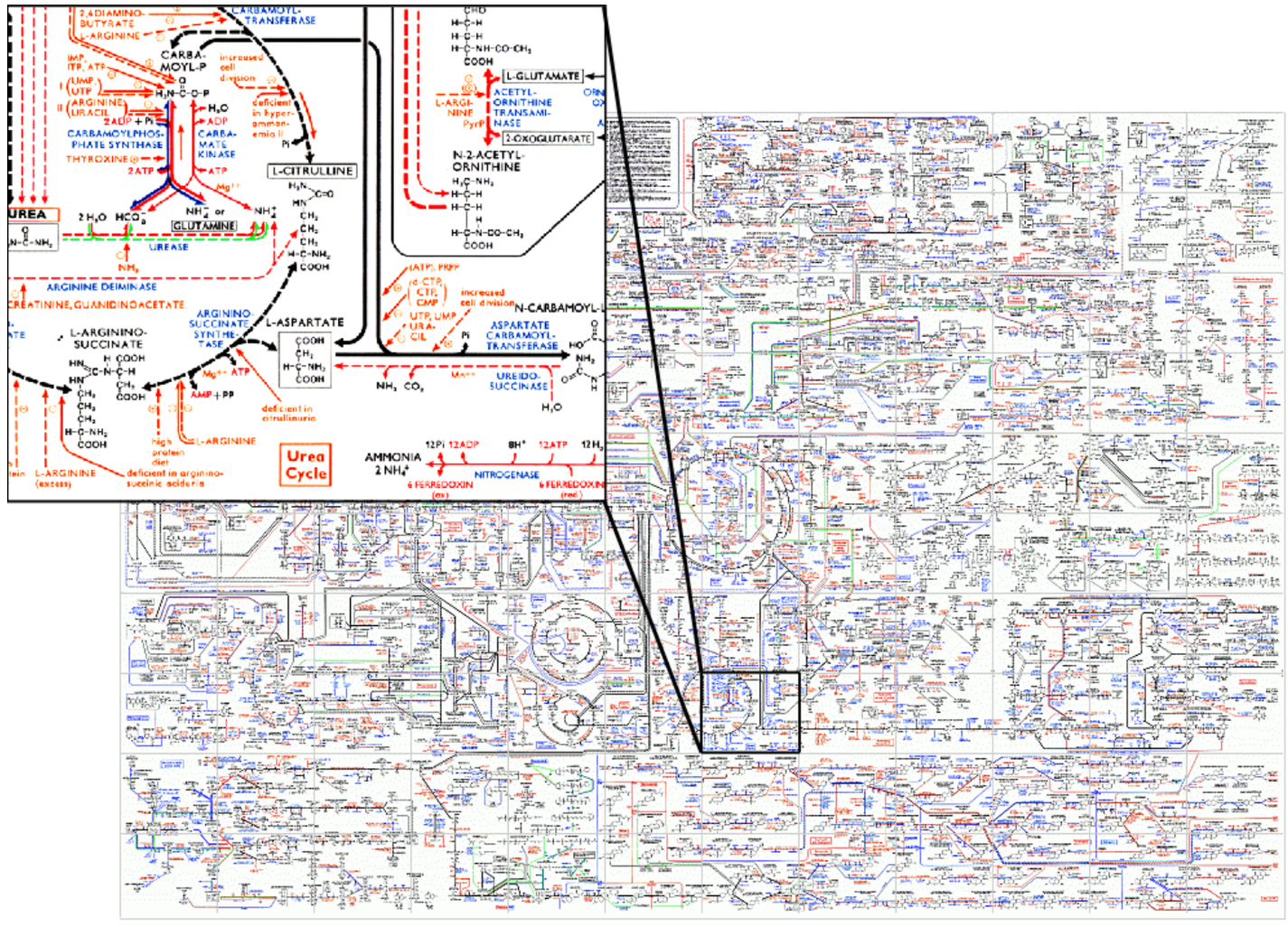
Data Analysis — Four steps

4/ **Bioinformatic analysis**: biological context, network analyses, data integration

(2) Reconstruction of networks from data

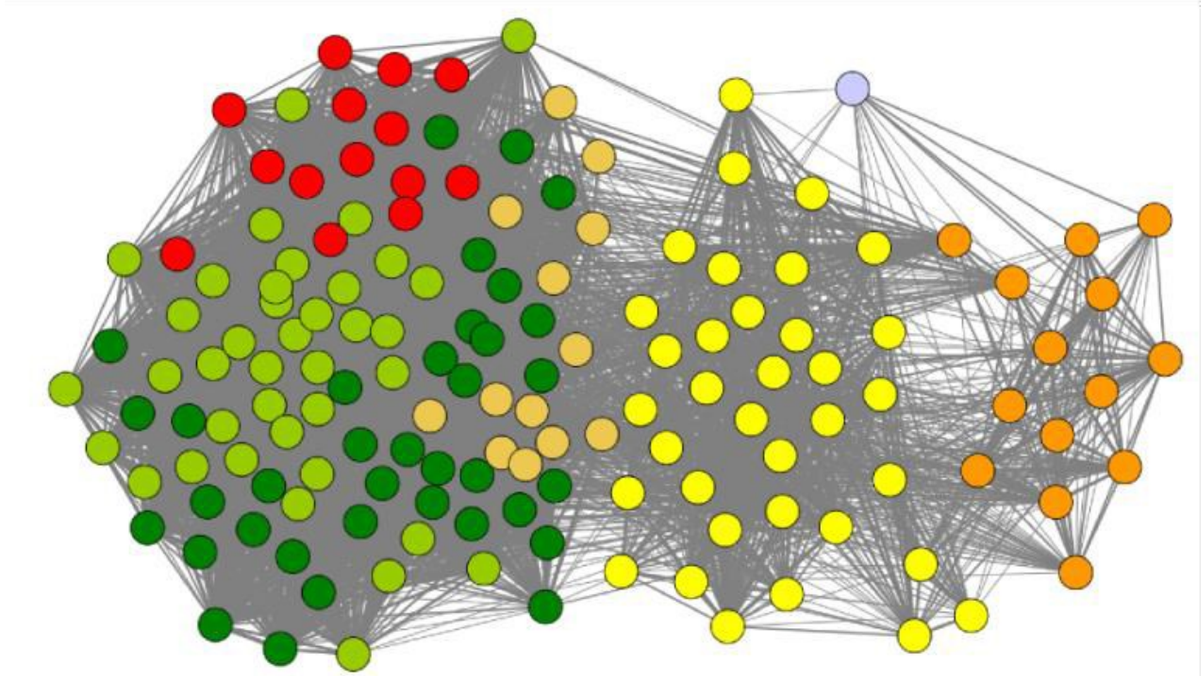
- **indirect effect**

—> corrected using partial correlation networks e.g. by Gaussian graphical model (GGM)

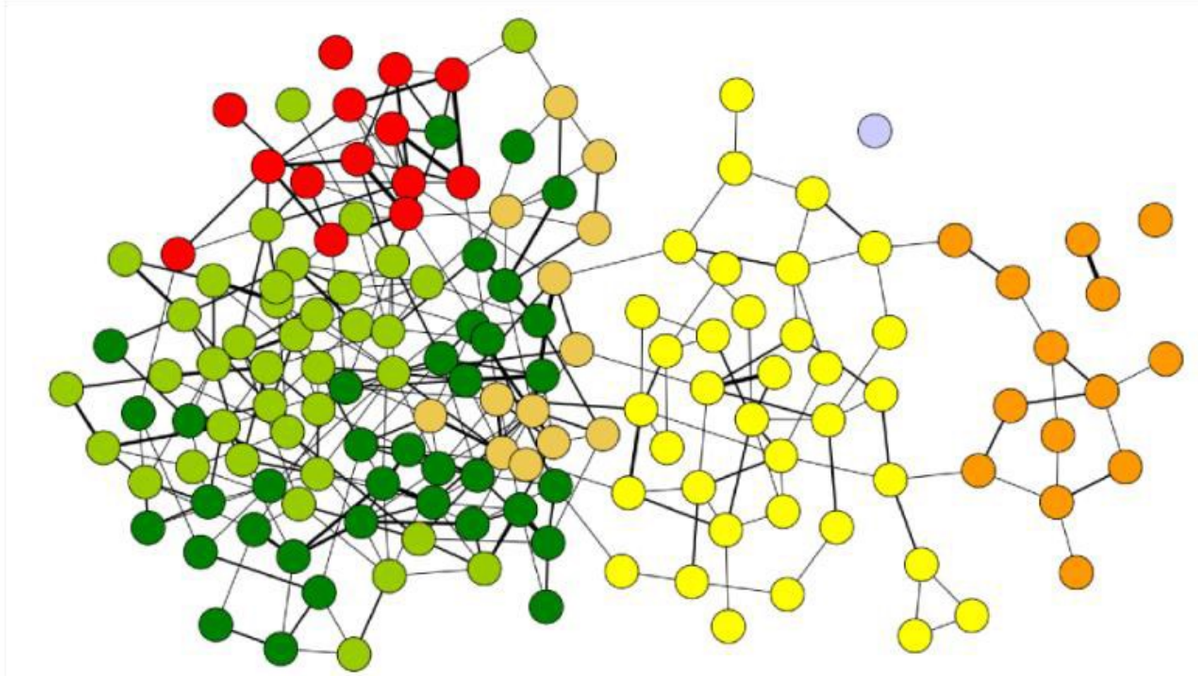


Reconstruction of metabolic networks using partial correlation networks (=GGMs)

Correlation network



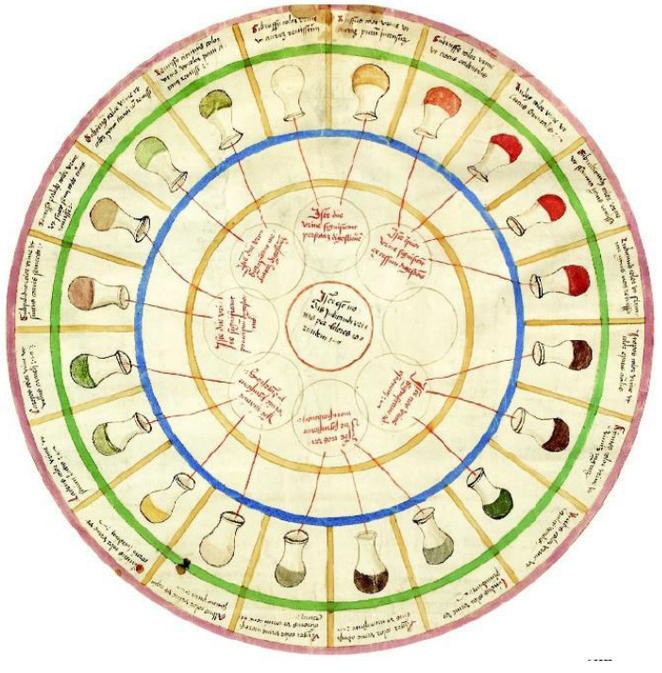
Gaussian graphical model



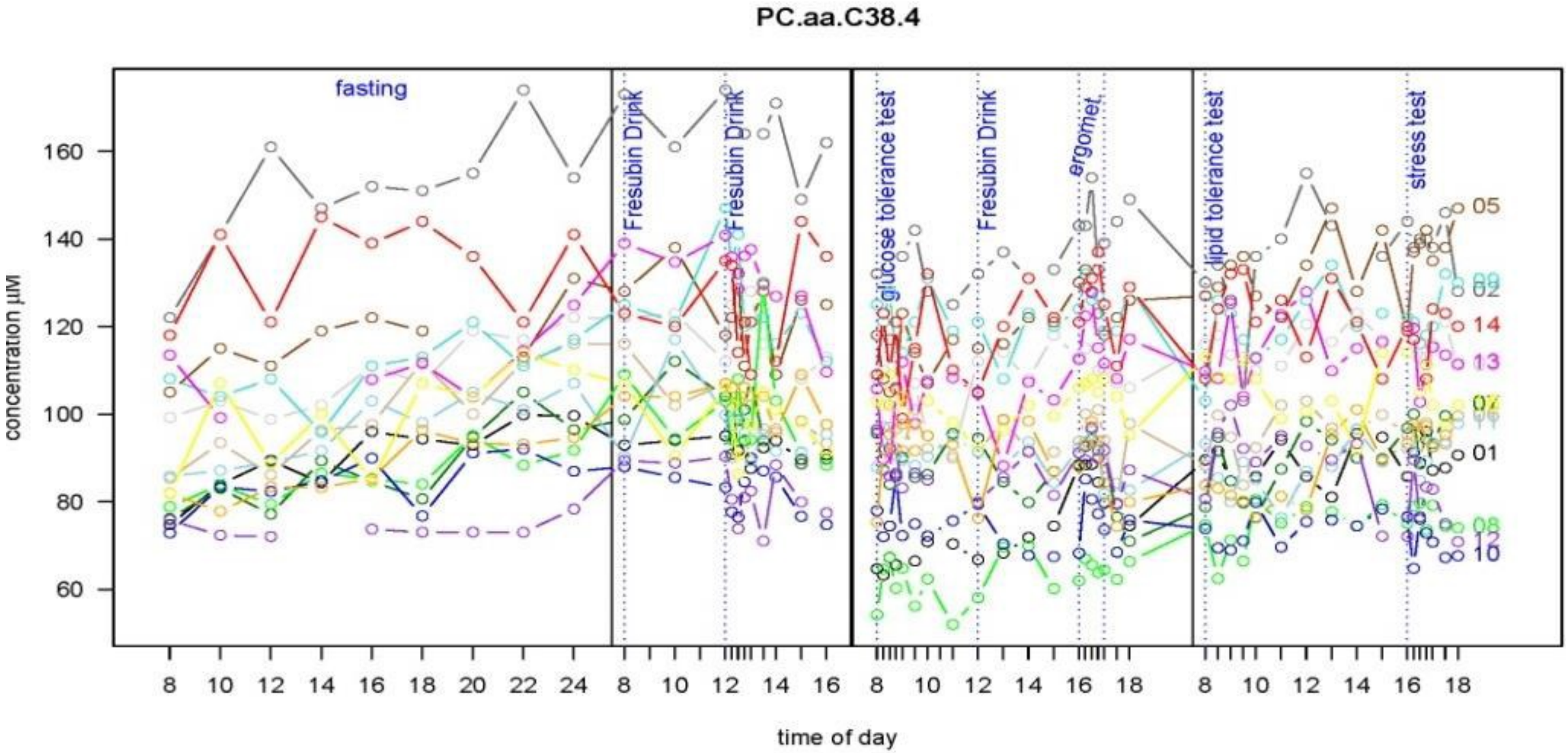
Applications and aims

- **„Biomarkers“ discovery**
  - Diagnosis (e.g. newborn screening)
  - Response on therapy
  - Stratification => Personalized Medicine

- **Pathomechanistic insights**
- **Preclinical drug testing**



Metabolic individuality

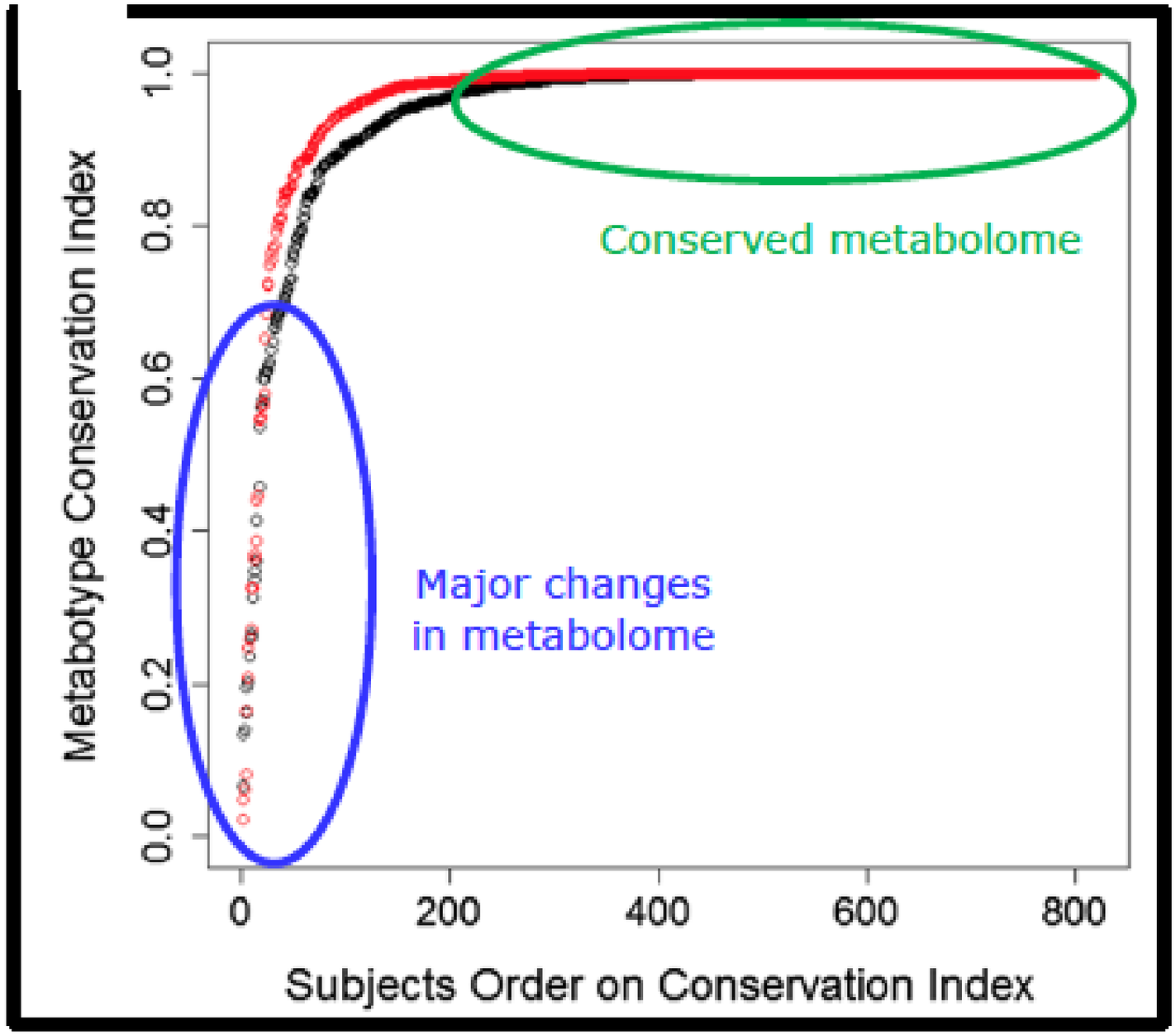


Long-term stability of metabolite profile

95% of subjects: "self" correlation ranked among the 30% strongest

Reasons for high conservation:

- **Genetic variation** (Heritability of metabolite levels)
- **Microbiome**
- **Lifestyle**



Yousri et al., Metabolomics, 2014