Metabolomics, an introduction with focus on data analysis

International Agency for Research on Cancer
Lyon, France

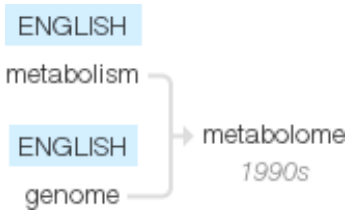
Reza Salek

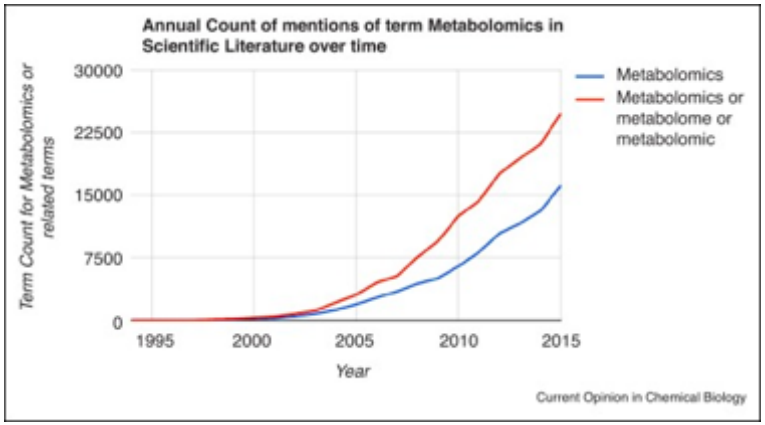
International Agency for Research on Cancer
World Health Organization

Some Definitions

metabolome

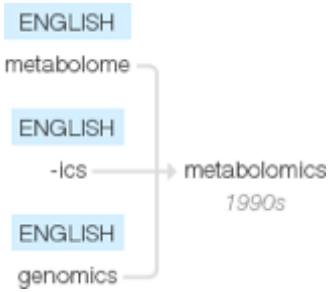
/mɪˈtabələʊm/
noun
BIOCHEMISTRY
noun: **metabolome; plural noun: metabolomes**
the total number of metabolites present within an organism, cell, or tissue.

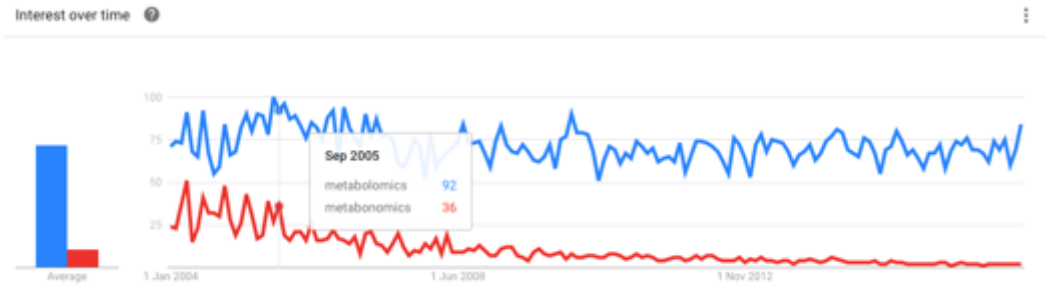




metabolomics

mɪˈtabəlɒmɪks/
Noun
BIOCHEMISTRY
noun: **metabolomics**
the scientific study of the set of metabolites present within an organism, cell, or tissue, often by measuring simultaneously (100s -10000s), many of which are not identified (features or analytes,)





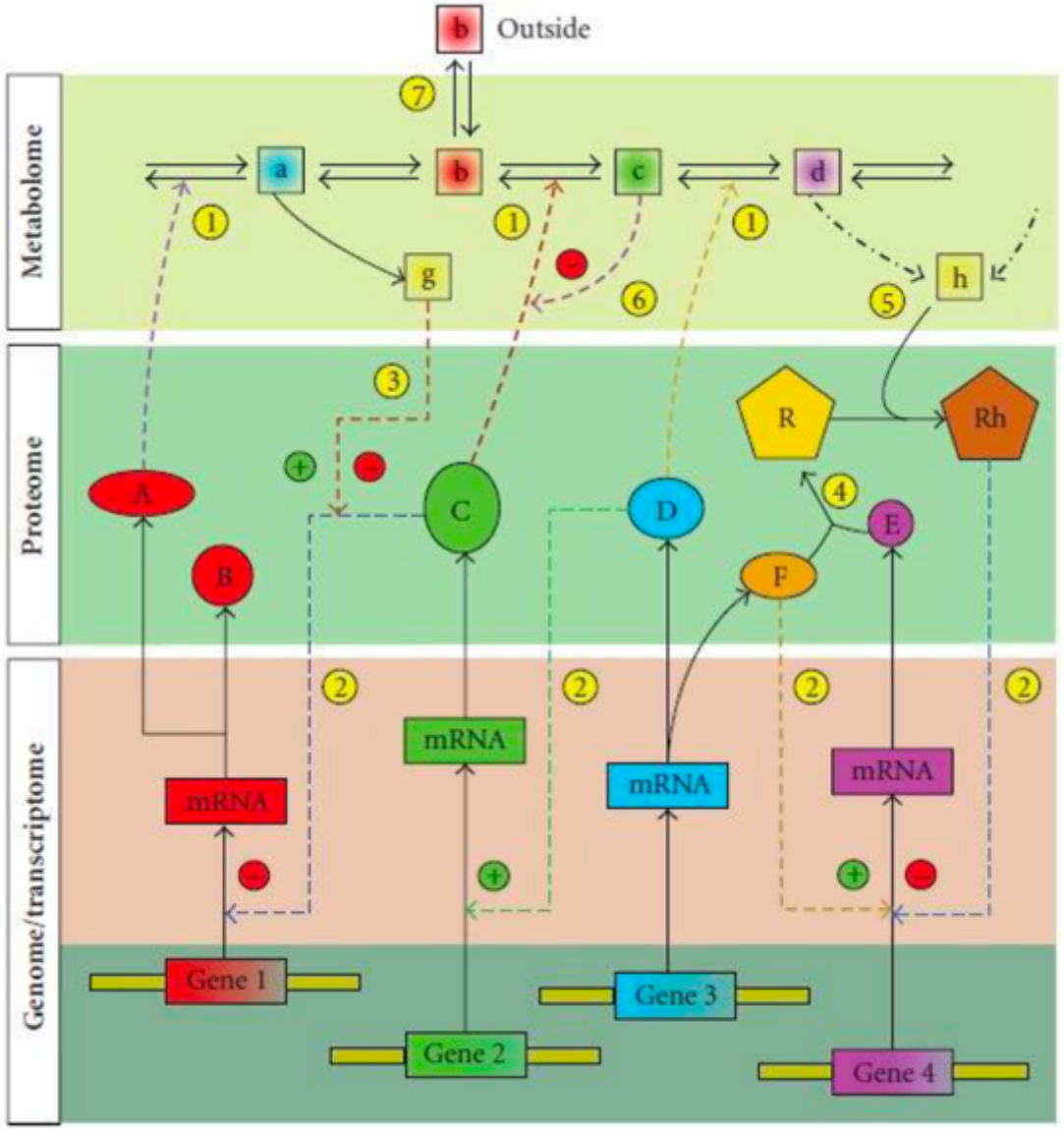
International Agency for Research on Cancer

World Health Organization

Genomics to metabolomics

- The system (cell) is more than the sum of its parts
- To understand the system we must study the system not the parts

- e.g. Silkworm metamorphosis



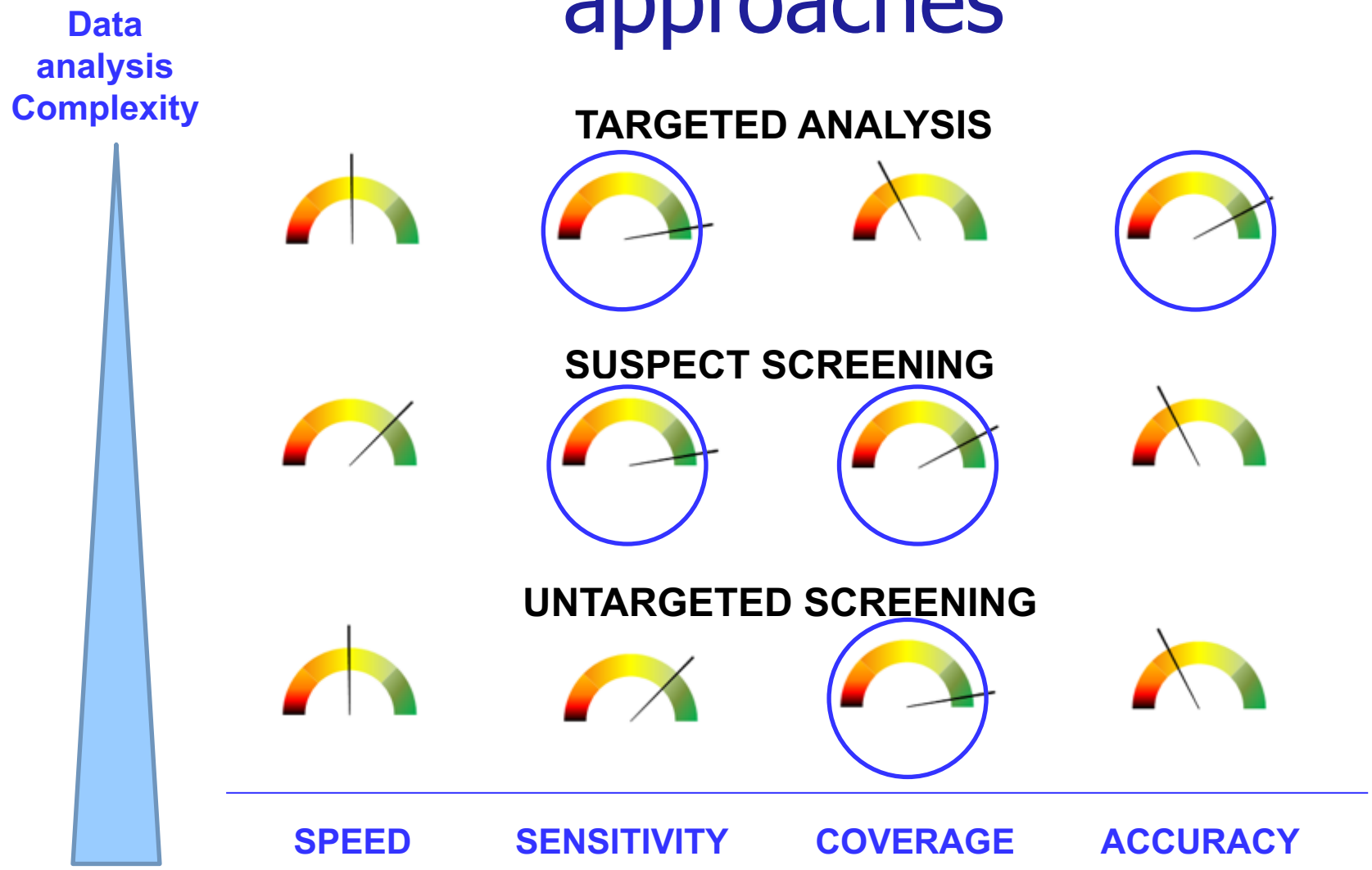
International Agency for Research on Cancer
World Health Organization
10.1155/2010/268925

Uroscopy – Early metabolomics

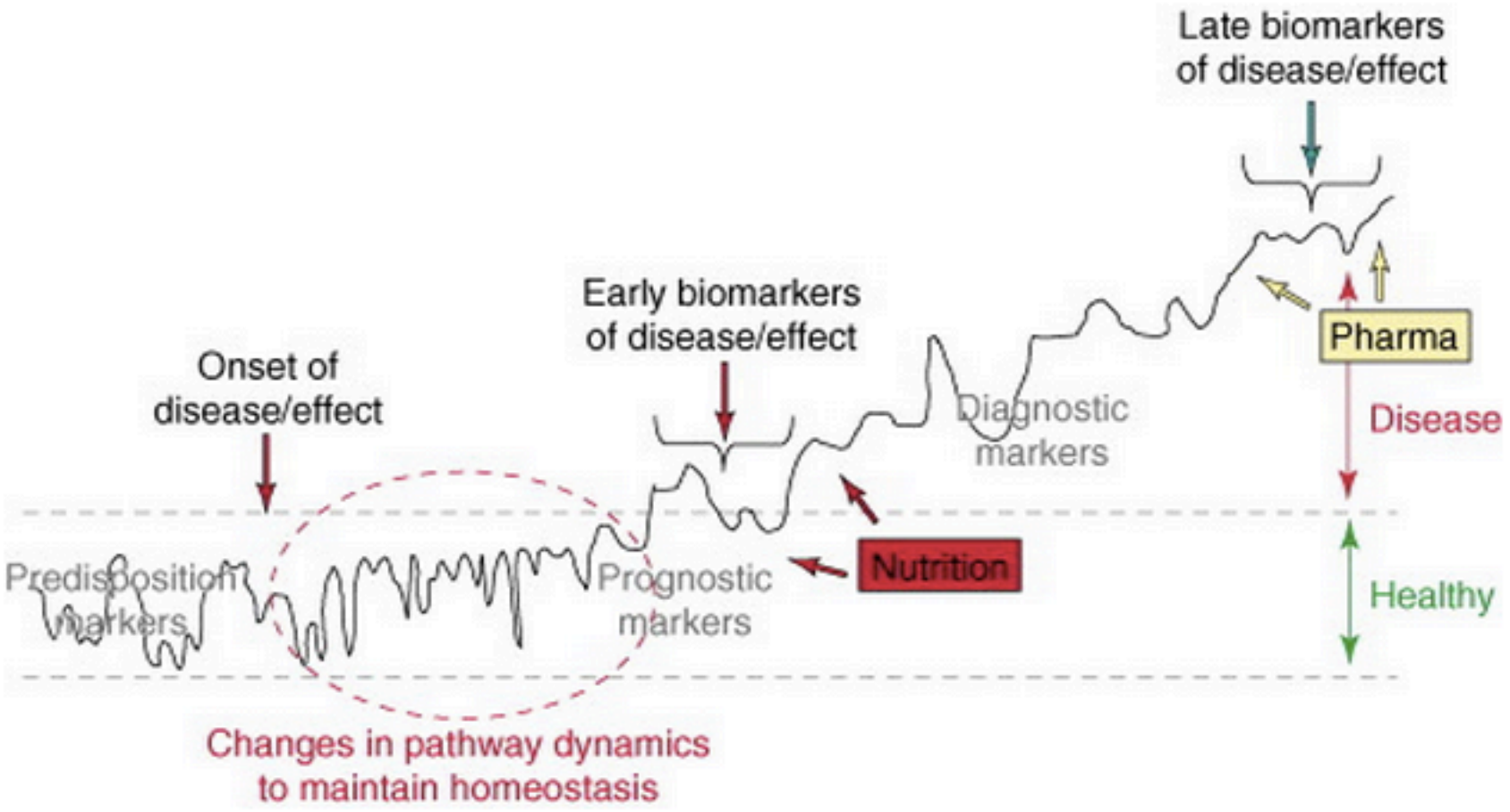
- It dates back to ancient Egypt, Babylon, and India. It was particularly emphasized in Byzantine medicine.
- The wheel describes the possible colors, smells and tastes of urine, and uses them to diagnose disease.

The Urine Wheel for diagnosing metabolic diseases, from *Epiphanie Medicorum* by Ullrich Pinder in 1506

Comparing different metabolomics approaches



Why is it important

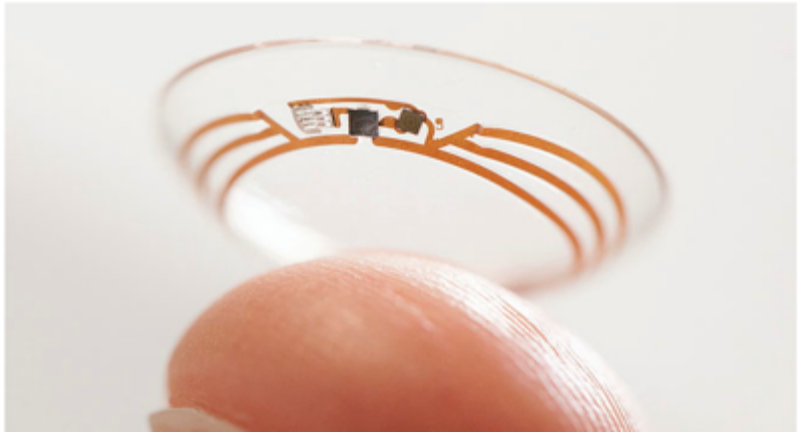


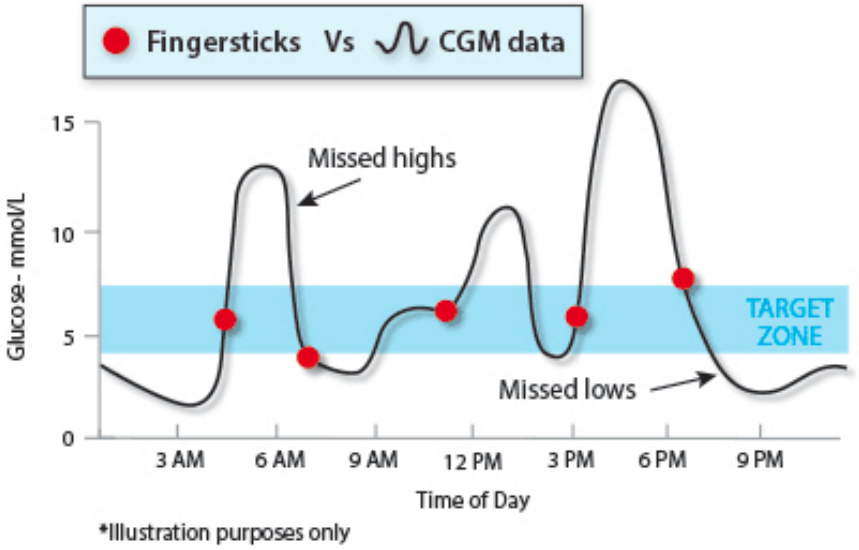
Jan van der Greef, Paul Stroobant, Rob van der Heijden, The role of analytical sciences in medical systems biology, Current Opinion in Chemical Biology, Volume 8, Issue 5, October 2004, Pages 559-565, ISSN 1367-5931, http://dx.doi.org/10.1016/j.cbpa.2004.08.013.

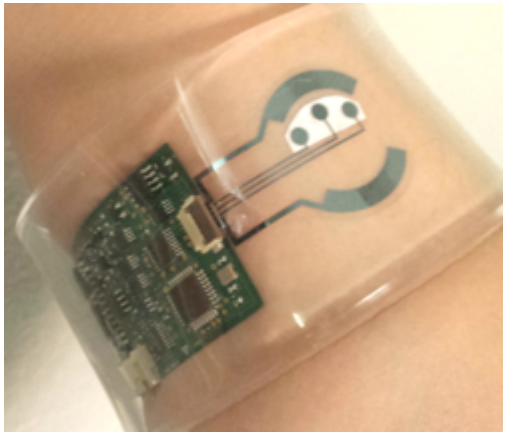
Bio Marker Diagnosis and Monitoring

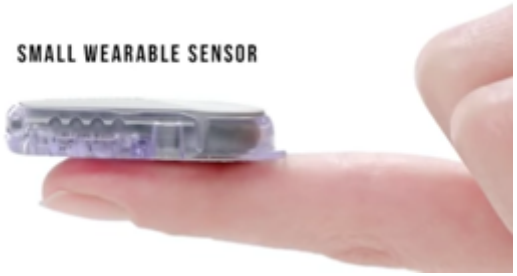
Google announces 'smart' contact lenses that monitor glucose levels

Published January 16, 2014 / FoxNews.com





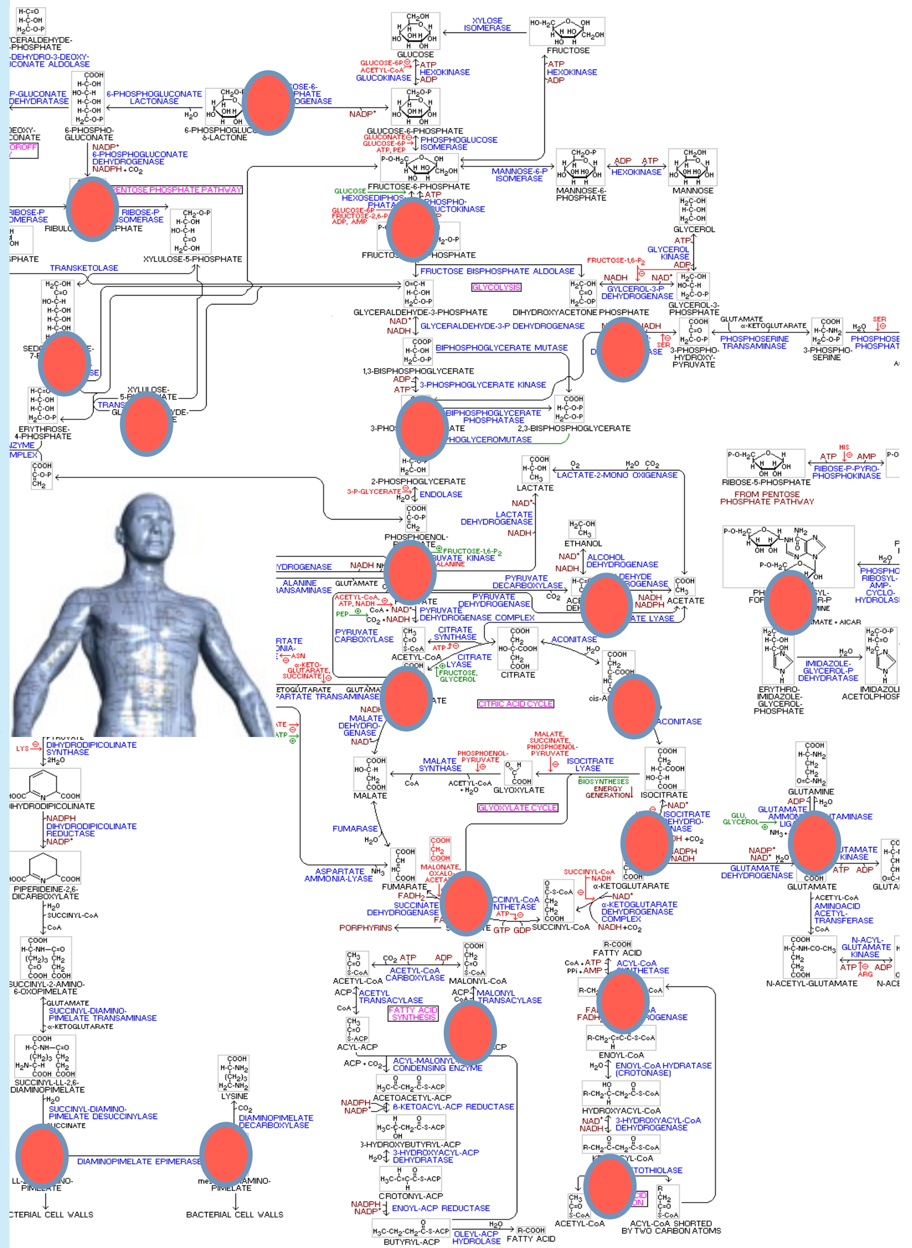




International Agency for Research on Cancer

World Health Organization


national rail
train operators
February 2012
PROJECT
MAPPING
LONDON
EDINBURGH
GLASGOW


GLUCOSE
FRUCTOSE
XYLOSE ISOMERASE
HEXOKINASE
GLUCOKINASE
GLUCOSE-6-PHOSPHATE
PHOSPHOGLUCOSE ISOMERASE
FRUCTOSE-6-PHOSPHATE
MANNOSE-6-P ISOMERASE
MANNOSE-6-PHOSPHATE
MANNOSE
GLYCEROL
GLYCEROL KINASE
6-PHOSPHOGLUCONATE LACTONASE
6-PHOSPHOGLUCONATE DEHYDROGENASE
PENTOSE PHOSPHATE PATHWAY
XYLULOSE-5-PHOSPHATE
TRANSKETOLASE
FRUCTOSE BISPHOSPHATE ALDOLASE
GLYCOLYSIS
GLYCERALDEHYDE-3-PHOSPHATE
DIHYDROXYACETONE PHOSPHATE
GLYCERALDEHYDE-3-P DEHYDROGENASE
BIPHOSPHOGLYCERATE MUTASE
1,3-BISPHOSPHOGLYCERATE
3-PHOSPHOGLYCERATE KINASE
2,3-BISPHOSPHOGLYCERATE
ERYTHROSE-4-PHOSPHATE
2-PHOSPHOGLYCERATE
ENOLASE
LACTATE
LACTATE-2-MONO OXIGENASE
LACTATE DEHYDROGENASE
ETHANOL
ALCOHOL DEHYDROGENASE
PYRUVATE DECARBOXYLASE
PYRUVATE DEHYDROGENASE COMPLEX
PYRUVATE CARBOXYLASE
ACETYL-CoA
CITRATE SYNTHASE
CITRATE
ACONITASE
CITRIC ACID CYCLE
MALATE
MALATE SYNTHASE
GLYOXYLATE
ISOCITRATE LYASE
GLYOXYLATE CYCLE
ISOCITRATE
FUMARASE
FUMARATE
SUCCINATE DEHYDROGENASE
α-KETOGLUTARATE
α-KETOGLUTARATE DEHYDROGENASE COMPLEX
GLUTAMATE DEHYDROGENASE
GLUTAMINE
GLUTAMATE
N-ACETYL-GLUTAMATE
RIBOSE-5-PHOSPHATE
FROM PENTOSE PHOSPHATE PATHWAY
ASPARTATE AMMONIA-LYASE
PORPHYRINS
ACETYL-CoA CARBOXYLASE
MALONYL-CoA
FATTY ACID SYNTHESIS
ACYL-MALONYL-ACP CONDENSING ENZYME
ACETOACETYL-ACP
β-KETOACYL-ACP REDUCTASE
3-HYDROXYBUTYRYL-ACP
CROTONYL-ACP
ENOYL-ACP REDUCTASE
BUTYRYL-ACP
OLEYL-ACP HYDROLASE
FATTY ACID
ACYL-CoA SYNTHETASE
ENOYL-CoA
ENOYL-CoA HYDRATASE (CROTONASE)
HYDROXYACYL-CoA
3-HYDROXYACYL-CoA DEHYDROGENASE
ACYL-CoA SHORTED BY TWO CARBON ATOMS
DIHYDRODIPICOLINATE SYNTHASE
DIHYDRODIPICOLINATE REDUCTASE
PIPERIDEINE-2,6-DICARBOXYLATE
SUCCINYL-DIAMINO-PIMELATE TRANSAMINASE
SUCCINYL-DIAMINO-PIMELATE DESUCCINYLASE
DIAMINOPIMELATE EPIMERASE
DIAMINOPIMELATE DECARBOXYLASE
LYSINE
BACTERIAL CELL WALLS

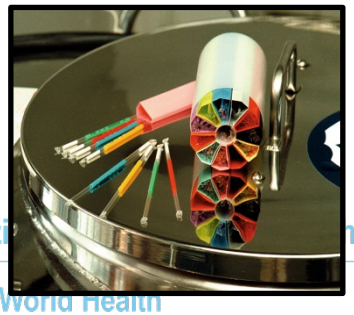
Part 1

WHAT WE DO AT IARC

EPIC: European Prospective Investigation into Cancer and nutrition

EPIC Project

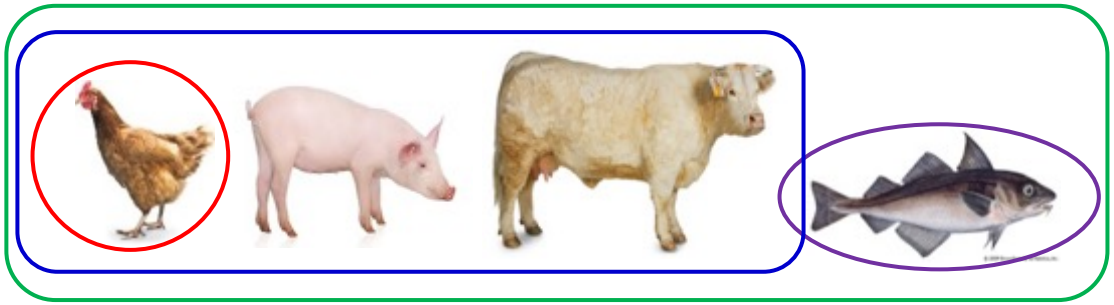
- Multicentric cohort study
- Nutrition, lifestyle and cancer
- 23 centres, 10 European countries
- 370.000 women, 150.000 men (aged 35-74) recruited between 1992-1998
- Standardized questionnaireson diet &lifestyle
- Anthropometry measured
- Blood samplers (30ml) on 65% women et 93% men :



8 straws serum
12 straws of plasma
4 straws of buffy coat
4 straws of red blood cells

EWAS and dietary intake - Animal foods

Dietary intervention study
4 parallel groups (n=10 per group)
Controlled diet
3 weeks with increasing level of meat/fish intake
24-hr urine, serum
UHPLC-QTof-MS



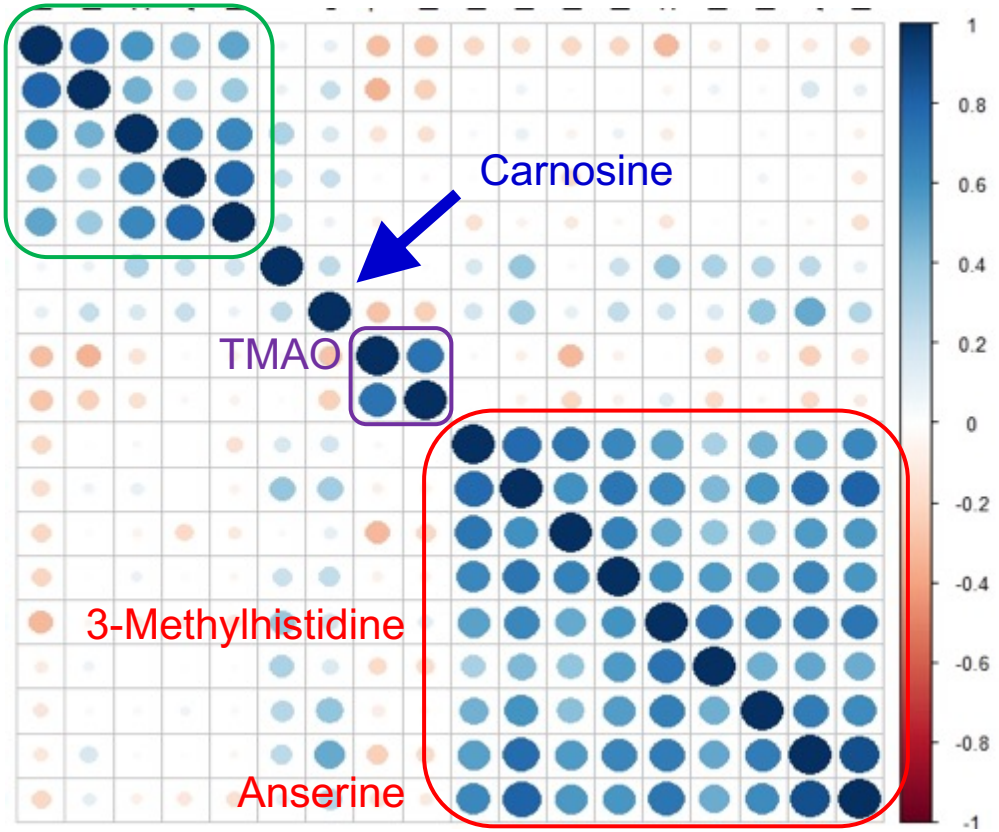
2-Methylbutyrylcarnitine
Acetylcarnitine
Propionylcarnitine

Carnosine

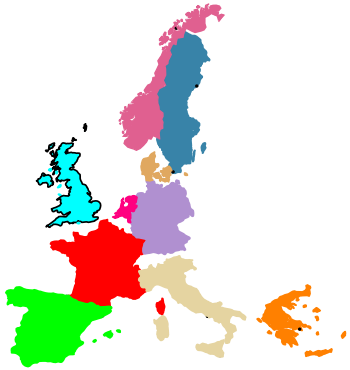
TMAO

3-Methylhistidine

Anserine



Observational study
24-hr dietary recalls
5 groups (n=10) according to meat/fish intake
24-hr urine



epic
European Prospective Investigation of Cancer

International Agency for Research on Cancer

World Health Organization

Cheung et al., 2017, *Am. J. Clin. Nutr.*

Untargeted metabolomics and risk of hepatocellular carcinoma

129 HCC cases
129 matched controls
Plasma samples
Untargeted metabolomics
(LC-QTof-MS)

9,602 MS features detected

↓

92 discriminant metabolites

↓

46 metabolites annotated

Top 16 HCC metabolites	LC-MS Method	Multivariable OR (95% CI)
Retinol	RP+	0.27 (0.16 - 0.48)
Dehydroepiandrosterone sulfate	HILIC-	0.35 (0.22 - 0.57)
Glycerophosphocholine	RP+	0.44 (0.28 - 0.71)
γ-carboxyethylhydroxychroman	RP+	0.56 (0.39 - 0.81)
Creatine	RP+	0.56 (0.37 - 0.83)
Tyrosine	RP+	**2.04 (1.30 - 3.20)**
N1-Acetylspermidine	HILIC+	**2.16 (1.38 - 3.37)**
Isatin	RP+	**2.28 (1.38 - 3.75)**
***p*-Hydroxyphenyllactic acid**	HILIC-	**2.77 (1.58 - 4.83)**
Sphingosine	RP+	**2.79 (1.66 - 4.71)**
L,L-Cyclo(leucylprolyl)	RP+	**3.25 (1.91 - 5.53)**
Glycochenodeoxycholic acid	RP+	**3.31 (1.99 - 5.51)**
Glycocholic acid	RP+	**4.07 (2.32 - 7.14)**
7-Methylguanine	HILIC+	**6.78 (3.24 - 14.18)**

- Many metabolites replicated in in the ATBC cohort (Finland)

Stepien et al., to be submitted

http://exposome-explorer.iarc.fr/

exposome explorer

International Agency for Research on Cancer
World Health Organization

exposome explorer

compounds, foods, cancers

Biomarker data | Classifications | Publications | Structure search | Downloads | About

Exposome-Explorer

First database dedicated to biomarkers of exposure to environmental risk factors for diseases.

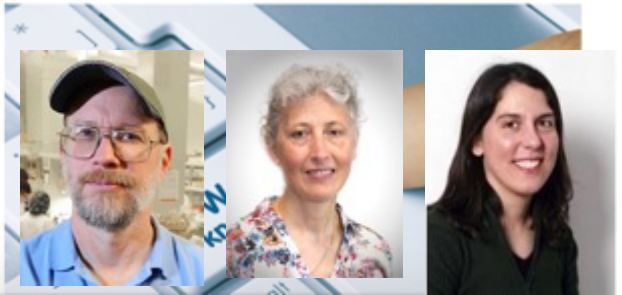
Read more

Neveu V, Nicolas G, Salek RM, Wishart DS, Scalbert A. Exposome-Explorer 2.0., 2019, *Nucl. Acid Res.*

Biomarker data

Classifications

Additional information

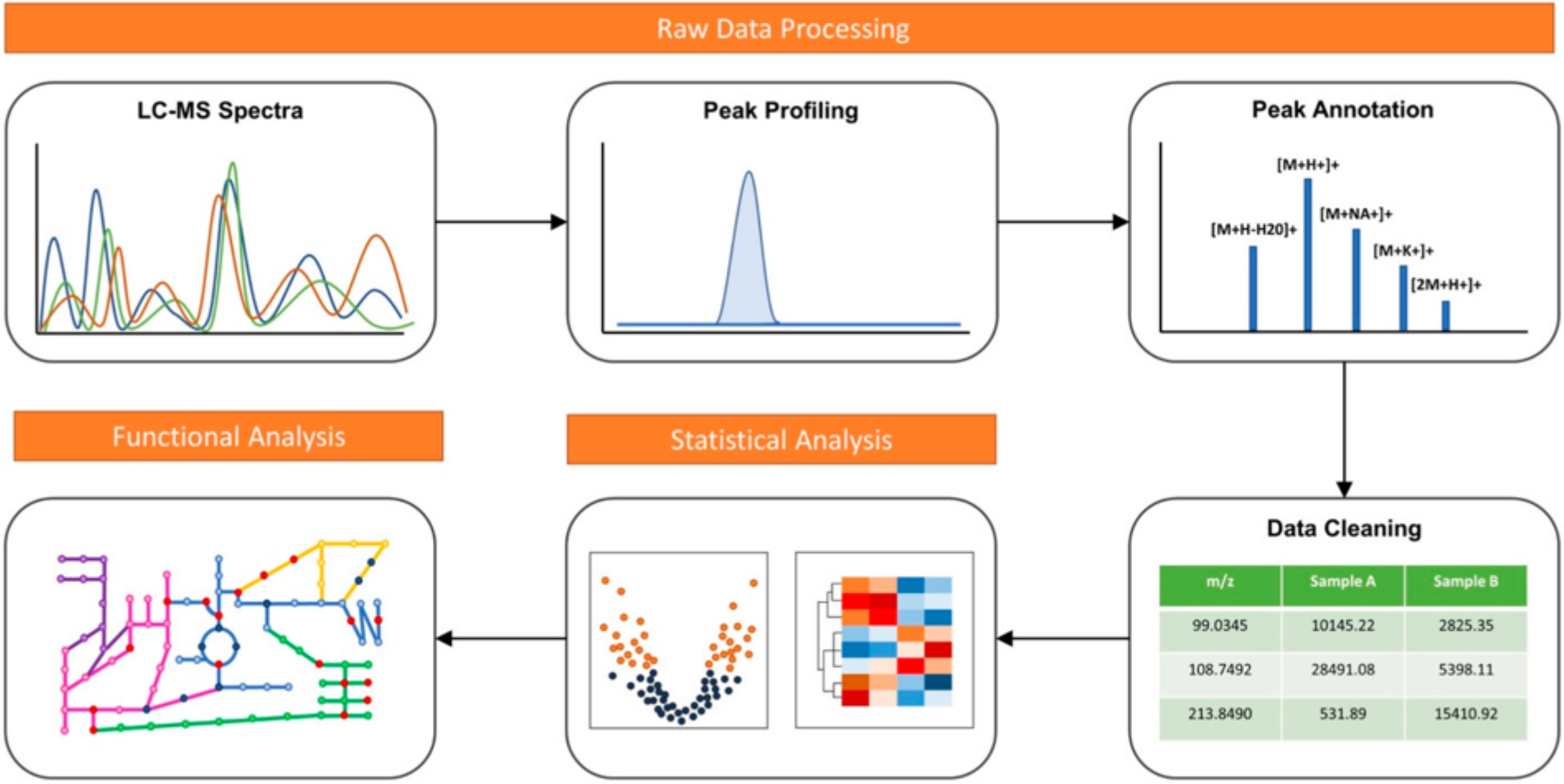


Raw Data to Peak tables

METABOLOMICS DATA

International Agency for Research on Cancer
World Health Organization

Overview of data handling in metabolomics

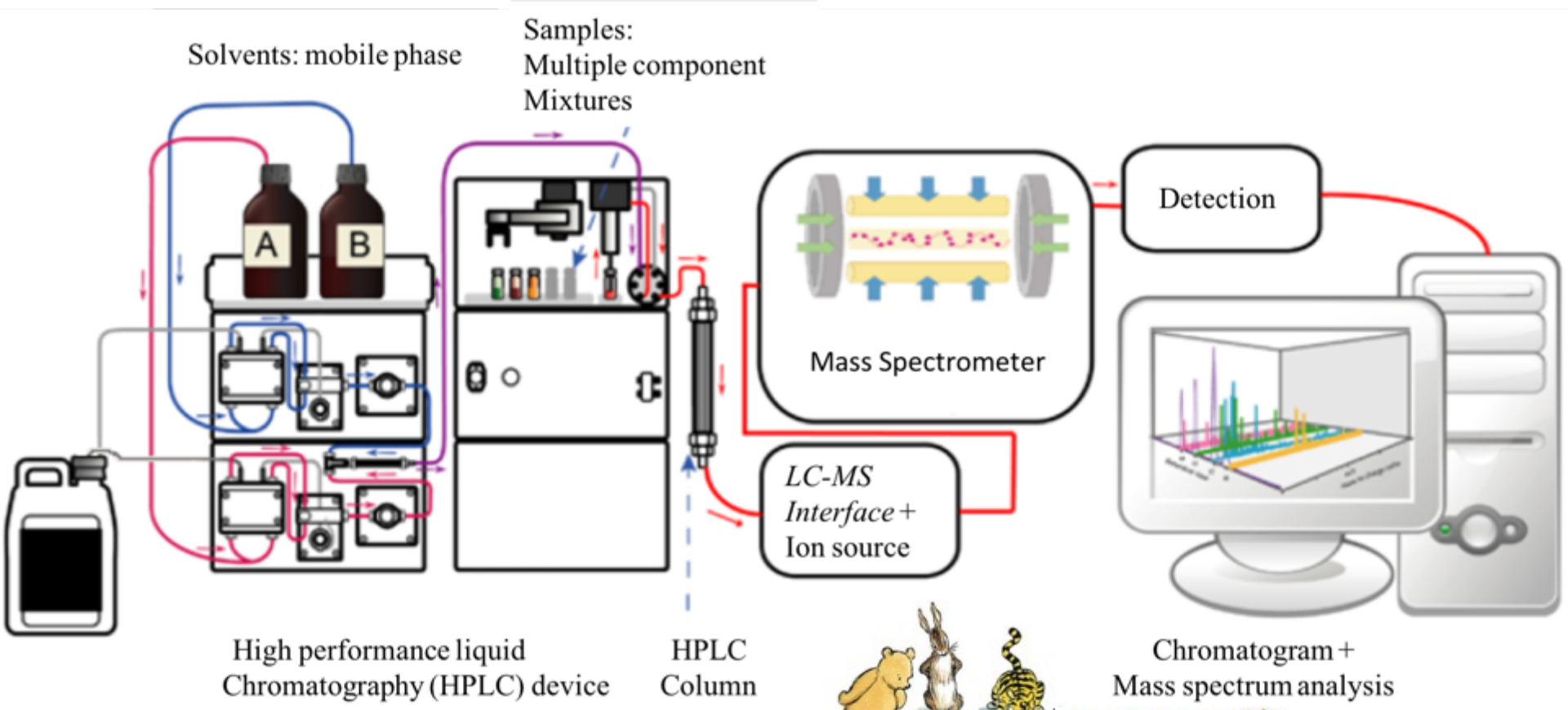


International Agency for Research on Cancer

World Health Organization

https://doi.org/10.3390/metabo9030057

Diagram of an LC-MS system

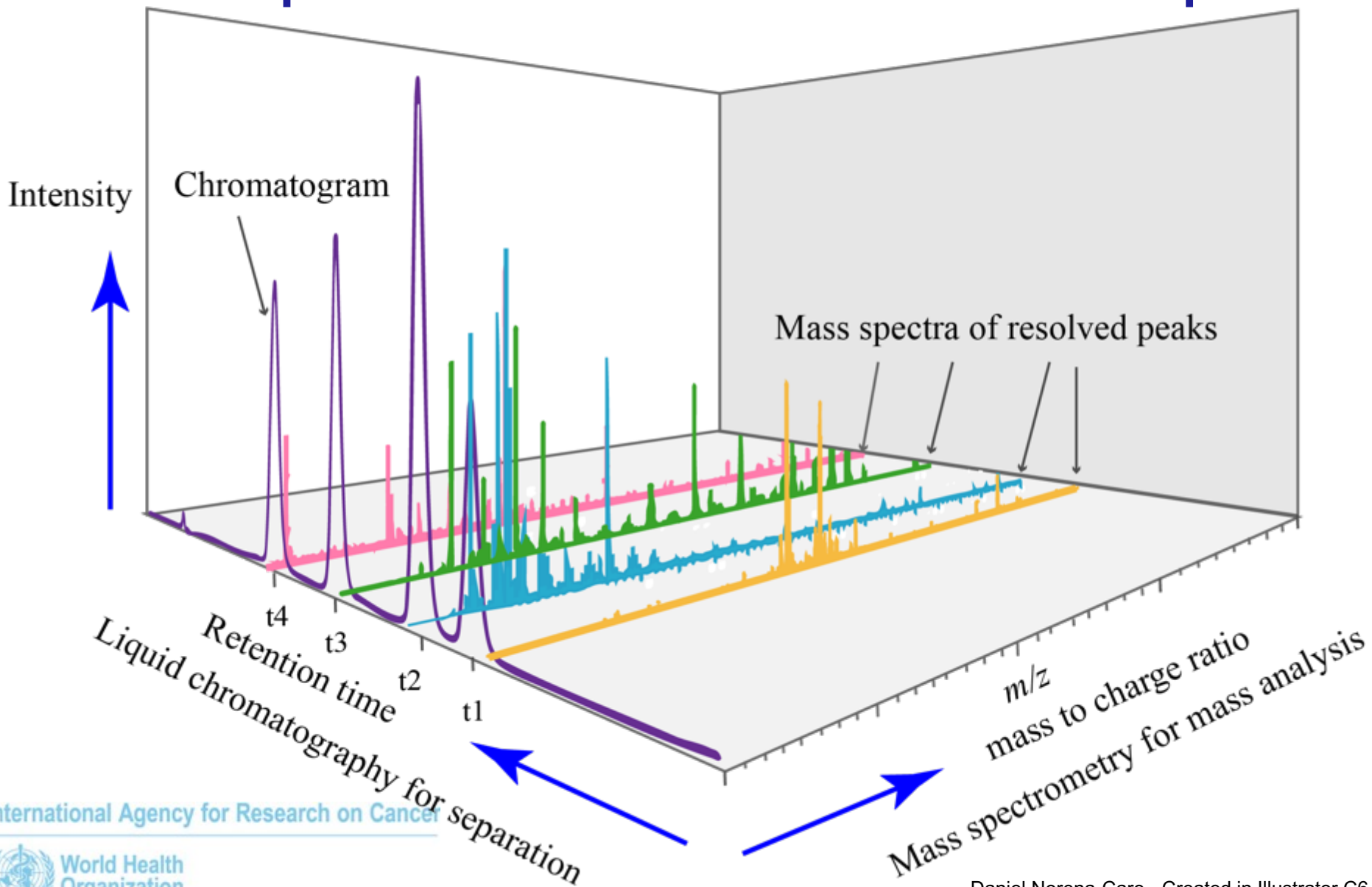


International Agency for Research on Cancer

World Health Organization

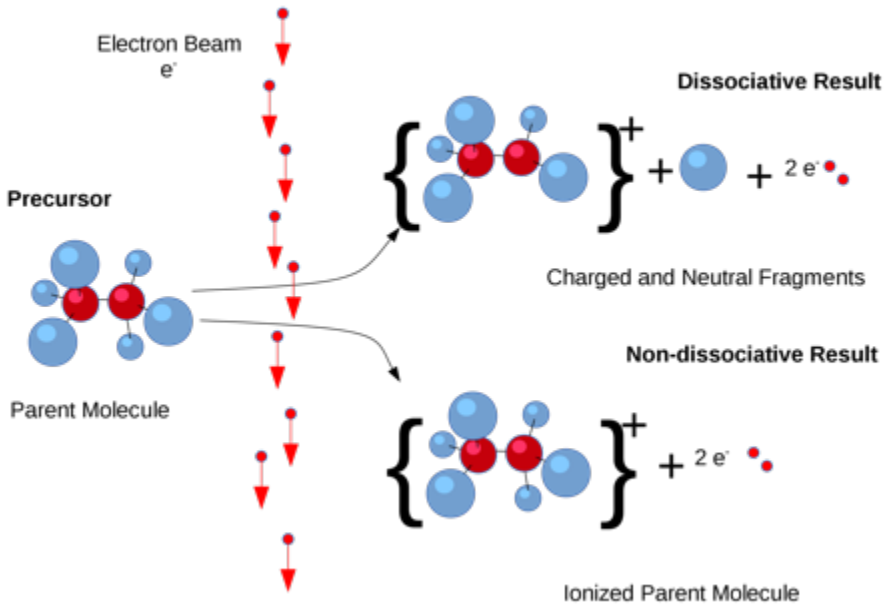
https://en.wikipedia.org/wiki/Liquid_chromatography–mass_spectrometry

LC-MS spectrum of each resolved peaK



International Agency for Research on Cancer
World Health Organization

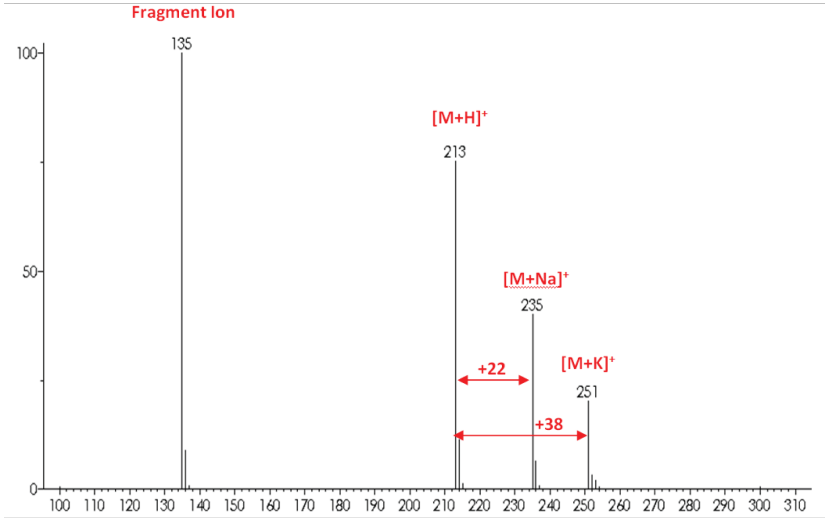
Daniel Norena-Caro - Created in Illustrator C6



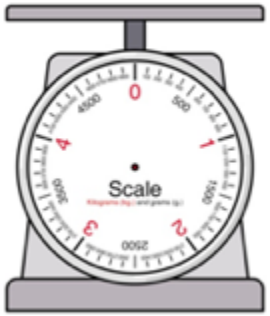
H1.0078 u - e 0.00054 u= 1.0073

Ionization	Formation	Ion Mass
Positive	$[M+H]^+$	m+1.0073
	$[M+2H]^{2+}$	m/2+1.0073
	$[M+Na]^+$	m+22.9892
	$[M+K]^+$	m+38.9632
	$[M+NH4]^+$	m+18.03382
Negative	$[M-H]^-$	m-1.0073
	$[M-2H]^{2-}$	m/2-1.0073
	$[M-2H+Na]^-$	m+20.9747
	$[M-2H+K]^-$	m+36.9486

M is the molecule with molecular weight m

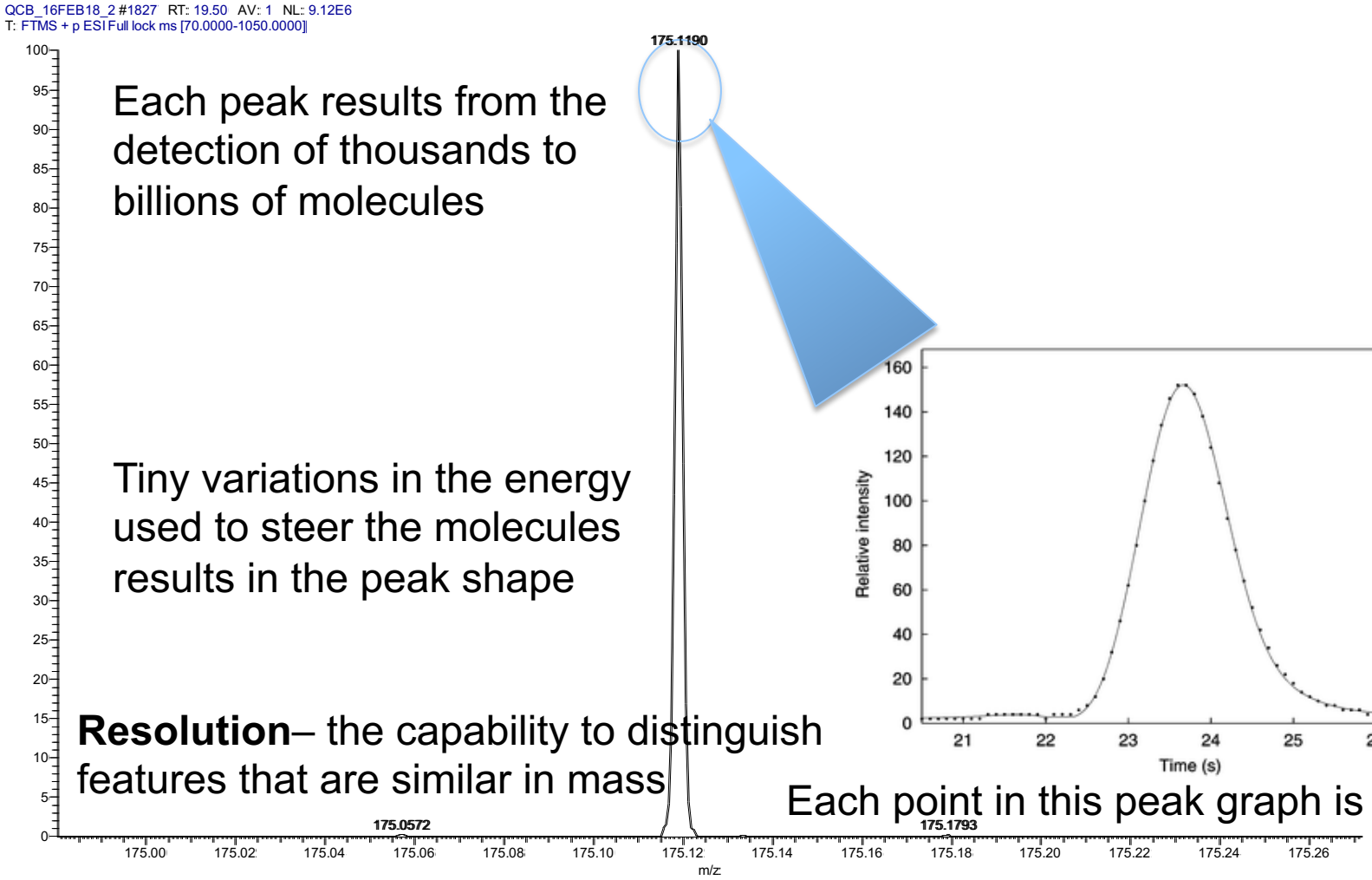


Molecules and fragments



Internat

MS Spectra

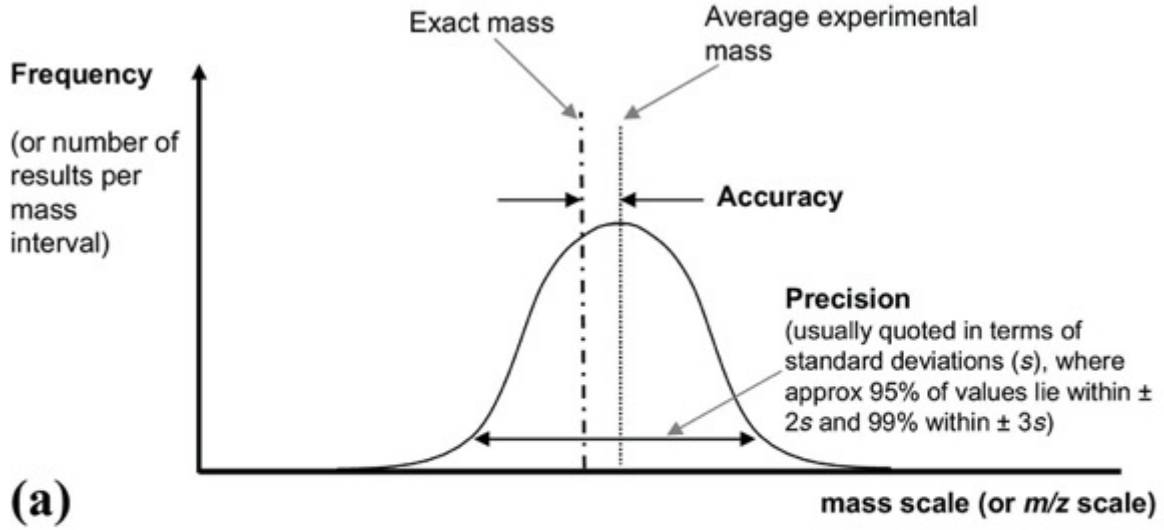


Each peak results from the detection of thousands to billions of molecules

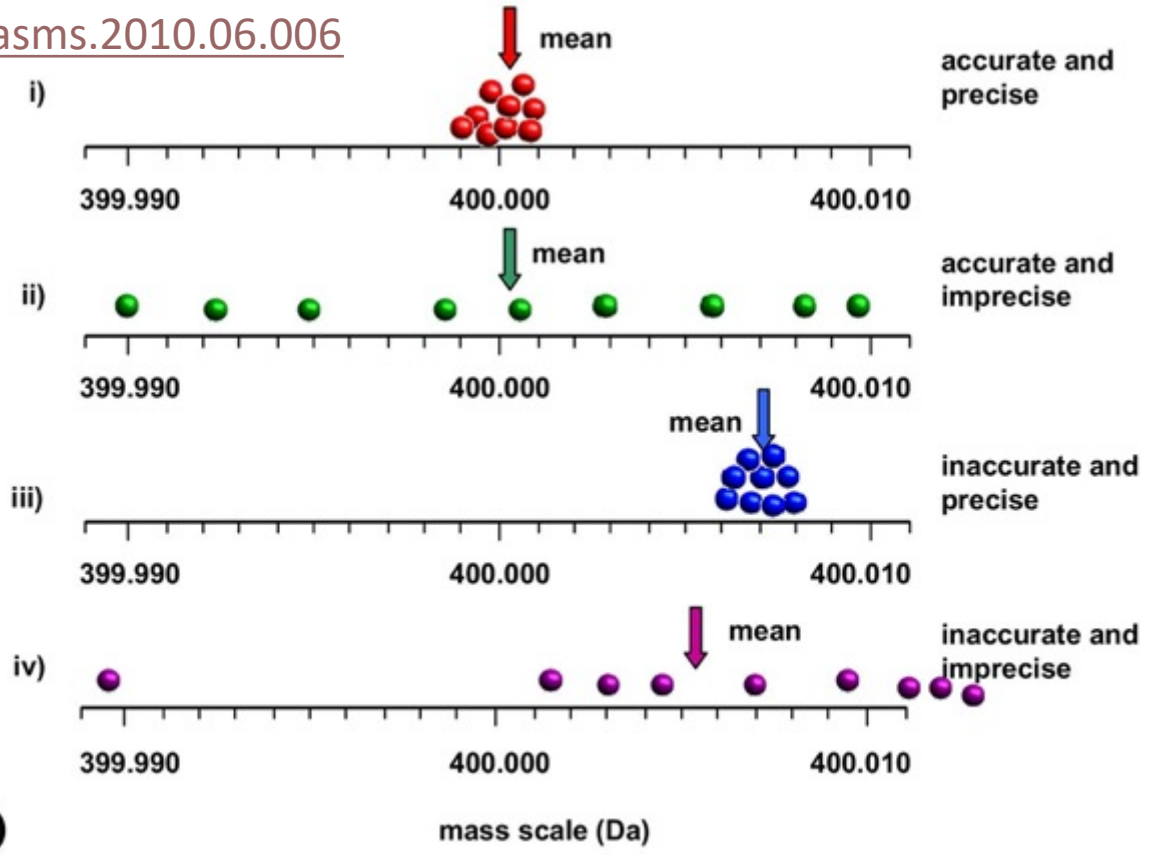
Tiny variations in the energy used to steer the molecules results in the peak shape

Resolution– the capability to distinguish features that are similar in mass

Each point in this peak graph is an MS scan



https://doi.org/10.1016/j.jasms.2010.06.006



International Agency for Research
World Health Organization

Exploiting High Mass Accuracy to ID Compounds

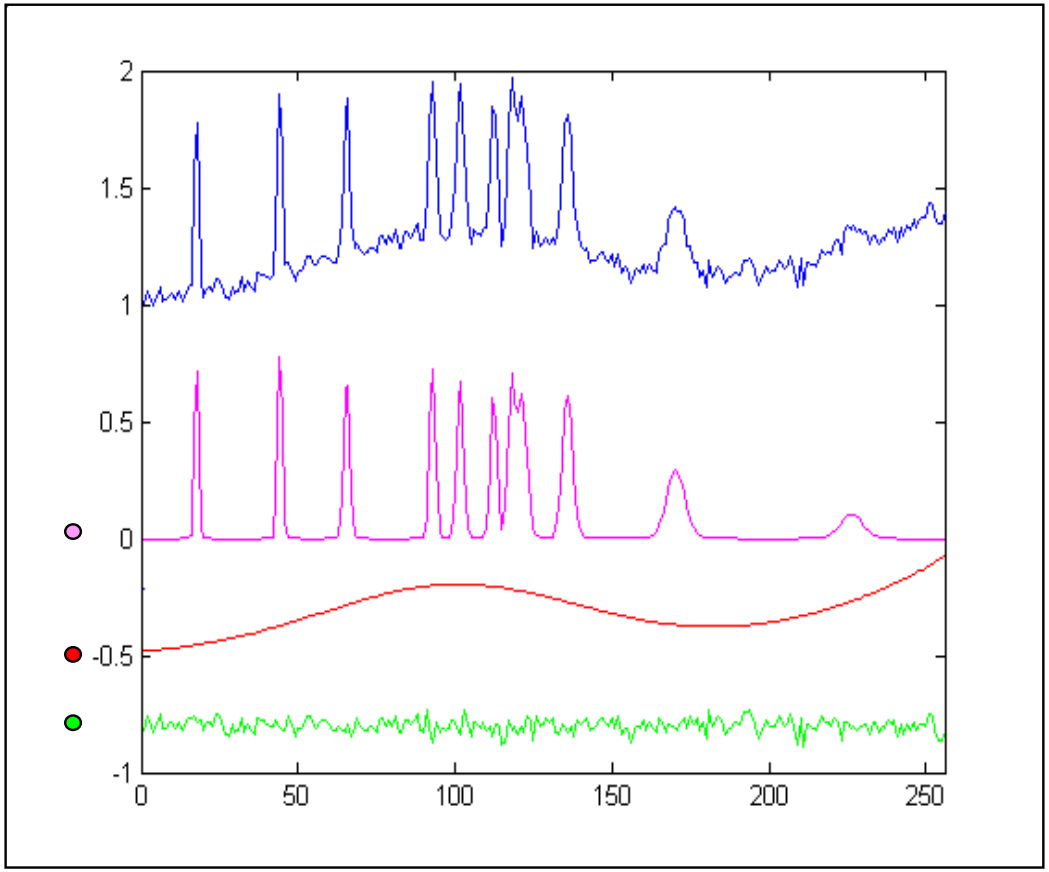
Type	Mass Accuracy
FT-ICR-MS	0.1 - 1 ppm
Orbitrap	0.5 - 1 ppm
Magnetic Sector	1 - 2 ppm
TOF-MS	3 - 5 ppm
Q-TOF	3 - 5 ppm
Triple Quad	3 - 5 ppm
Linear IonTrap	50-200 ppm (10 ppm in Ultra-Zoom)

$$ppm = \left(\frac{m_{exp} - m_{calc}}{m_{exp}}\right) * 1E+6$$

https://fiehnlab.ucdavis.edu/projects/seven-golden-rules/accurate-mass

Signal/Spectra components

- signal
- noise
- background



Signal enhancement

De-noising
Background correction

Example: XCMS Centwave algorithm

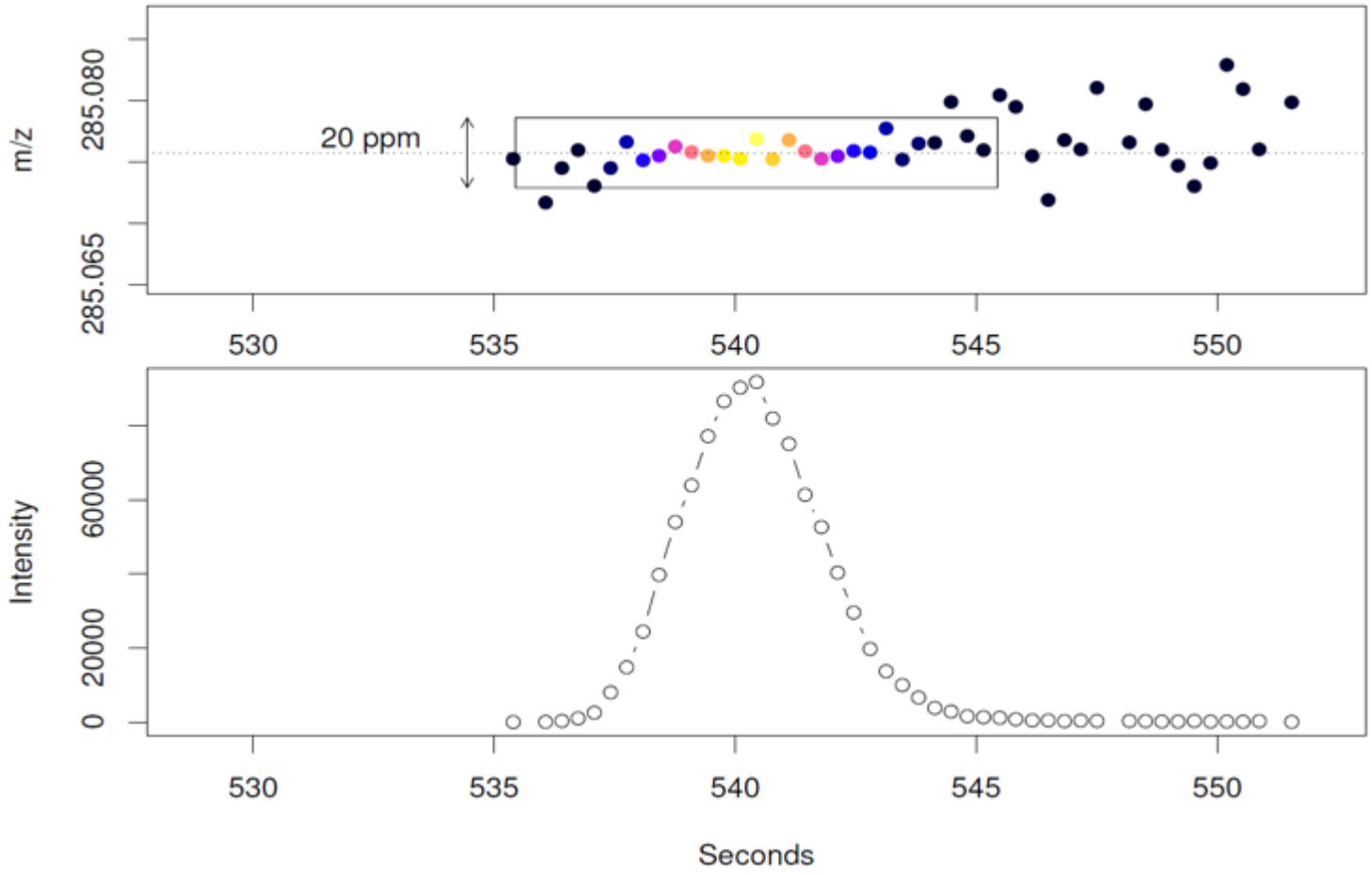


Figure from Tautenhahn, Böttcher and Neumann, 2008 (centWave paper)

XCMS centwave algorithm

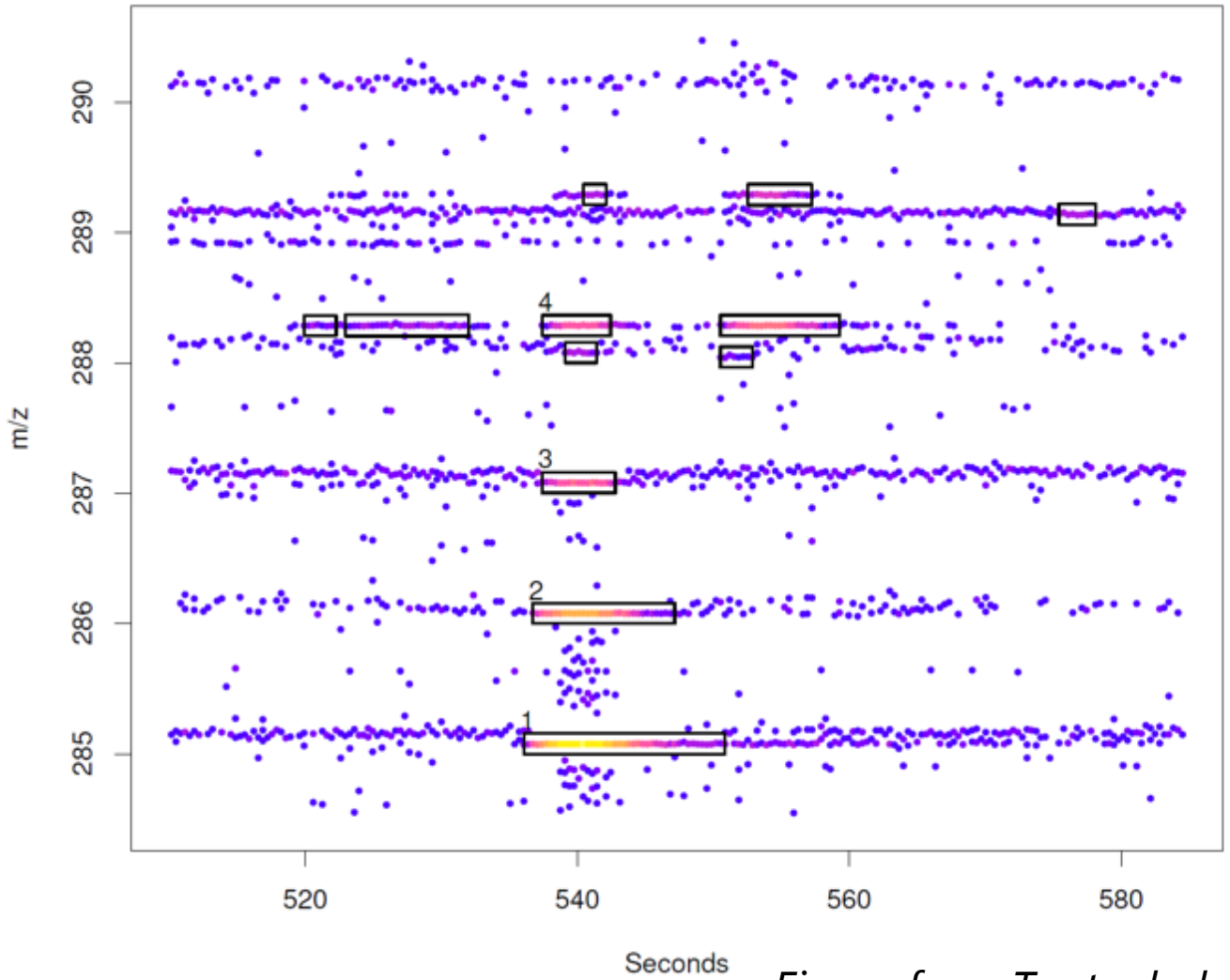
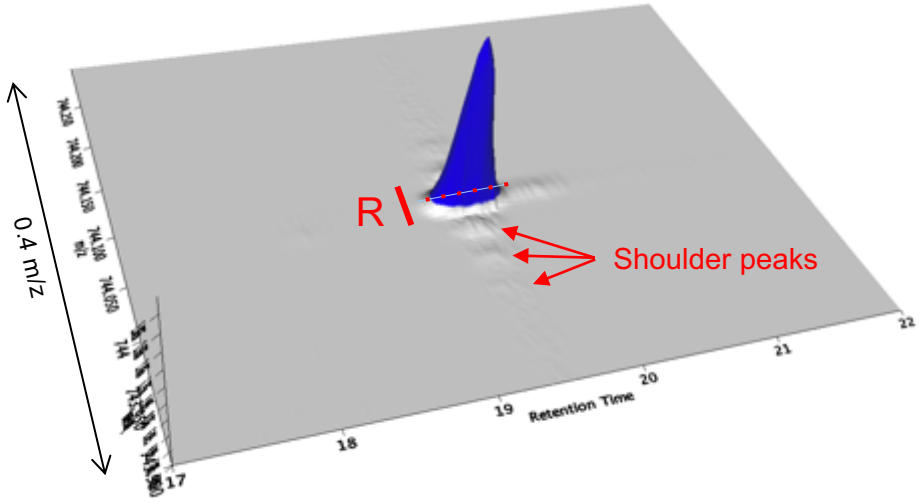


Figure from Tautenhahn, Böttcher and Neumann, 2008 (centWave paper)

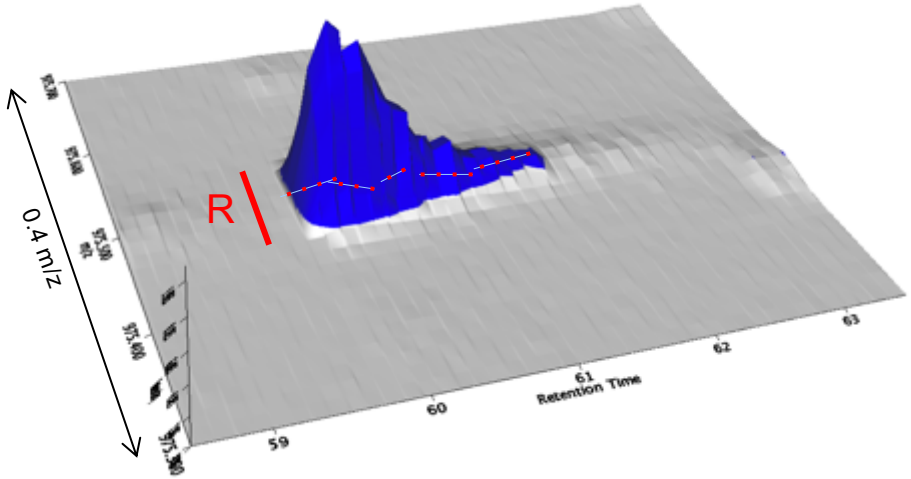
International A
World Health Organization

A single ion seen on different MS instruments

Thermo Orbitrap

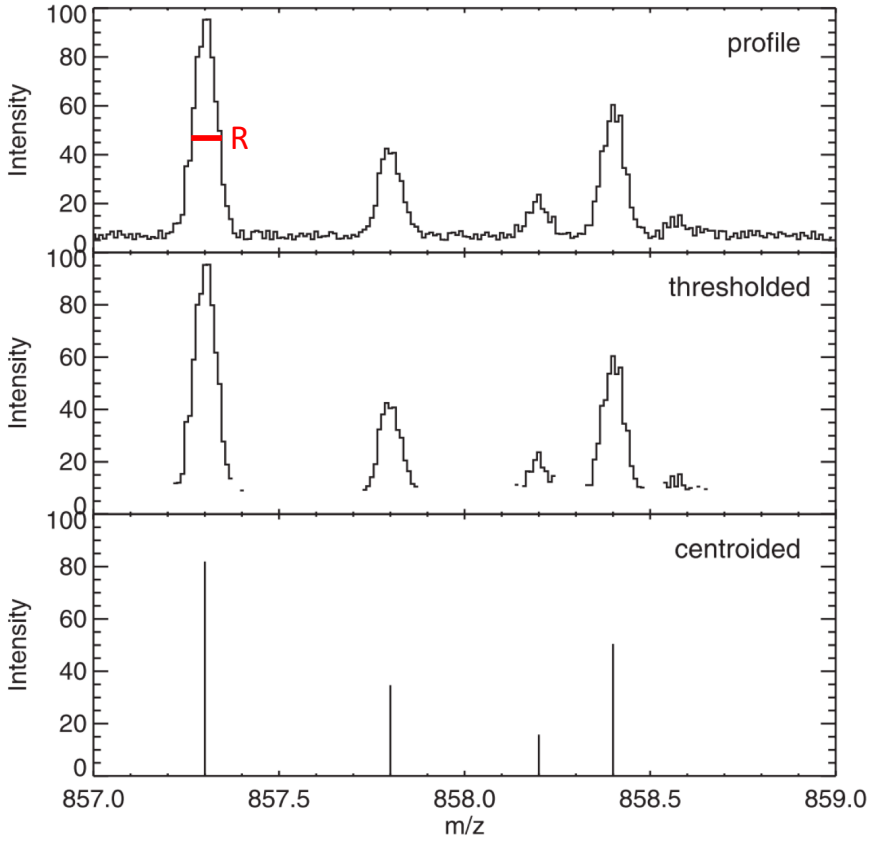


Waters SYNAPT G2-S



Slides by Tomáš Pluskal

Profile mode vs Centroid



Deutsch, *Mol. Cell. Proteomics* 2012

Terminology artifacts

Feature
Peak
Component
Isotope trace
Isotope envelope

Picking
Finding
Detection
Extraction
Deconvolution

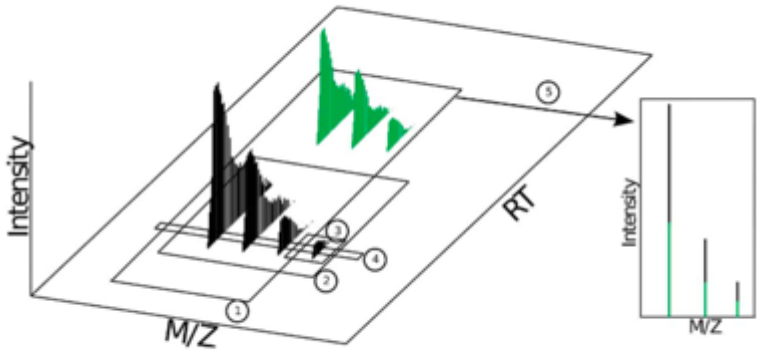


Figure 1 Any of the boxed signal in this figure can correctly be called a *chromatogram*, a *mass spectrum*, or a *profile spectrum* as defined by existing CVs. Additionally, the IUPAC terms *isotope pattern* and *isotope cluster* can refer to the signal in box 1, 2, or 3. It is clear that these terms pertain to too many distinct concepts to be clear. Note that existing nomenclatures have no way of distinguishing between these distinct data concepts, which can include parts of one or more combined signals.

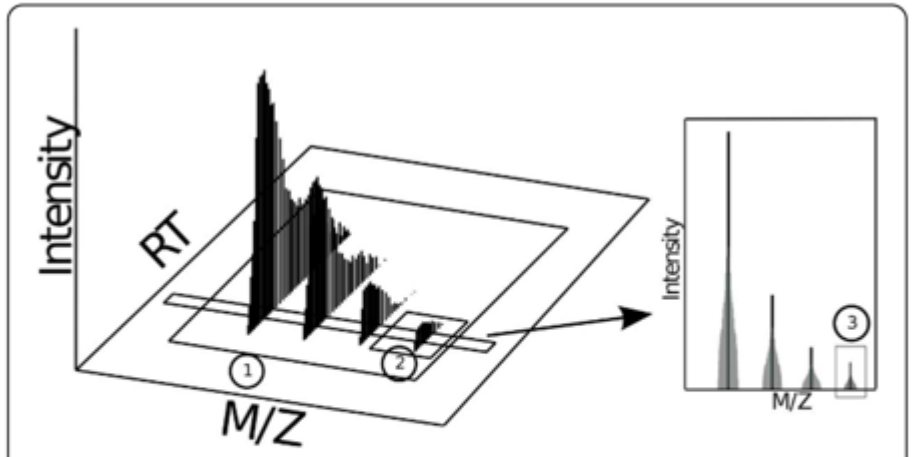
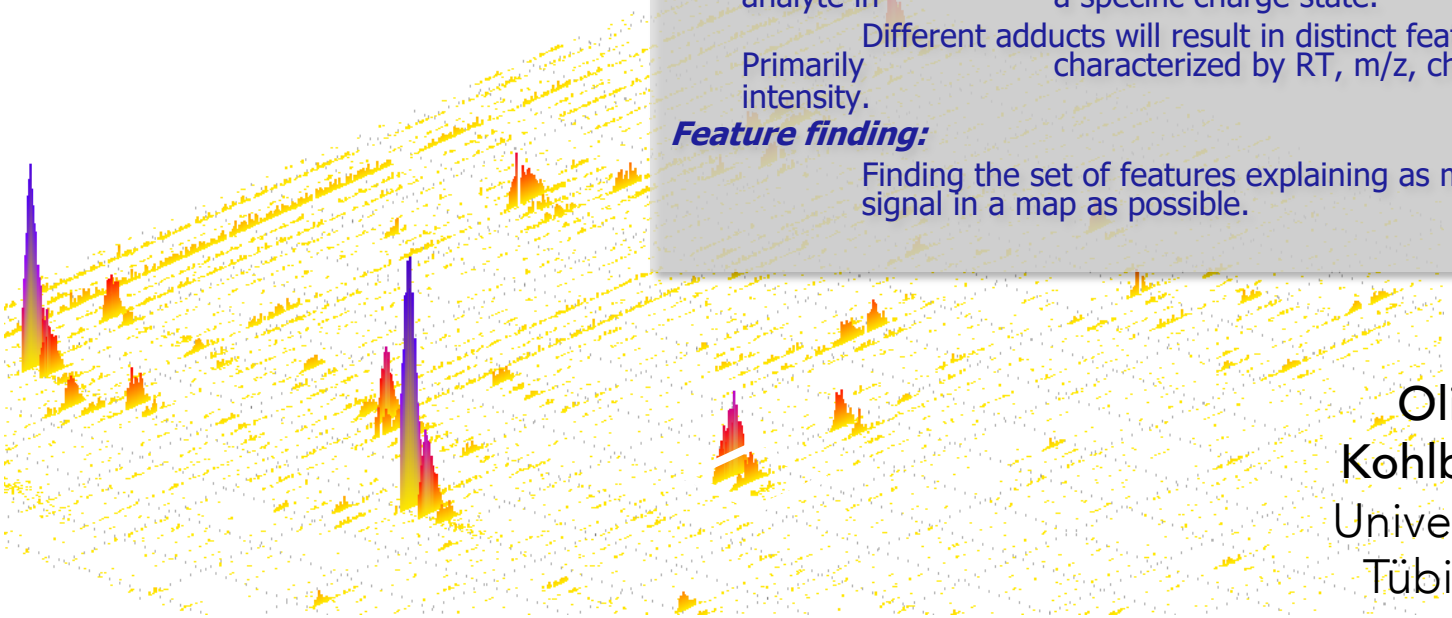


Figure 2 The PSI CV states that a *peak* is "A localized region of relatively large ion signal in a mass spectrum." Note that each of the numbered data segmentations in this figure qualify as a *peak* according to the PSI CV. In the literature, each of the illustrated concepts can also be found referred to as a *feature*.

Smith, R., Taylor, R. M. & Prince, J. T. Current controlled vocabularies are insufficient to uniquely map molecular entities to mass spectrometry signal. *BMC Bioinformatics* 16 Suppl 7, S2 (2015)

Feature Finding – Terms



Map:

Two-dimensional data set (RT, m/z) containing the MS signal from one LC-MS run.

Feature:

The sum of all the MS signals caused by the same analyte in a specific charge state.

Different adducts will result in distinct features. Primarily characterized by RT, m/z, charge, intensity.

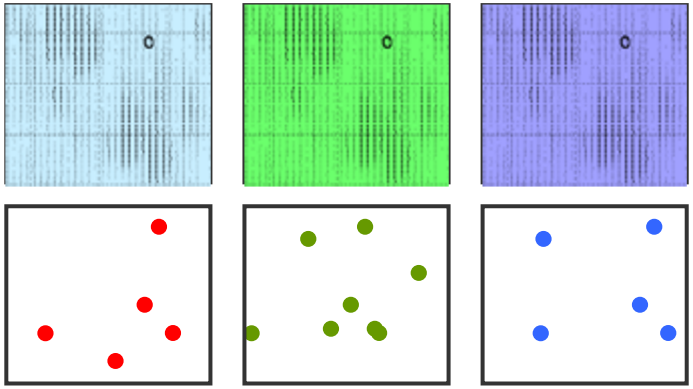
Feature finding:

Finding the set of features explaining as much of the signal in a map as possible.

Oliver Kohlbacher
University of Tübingen

Metabolic Profiling

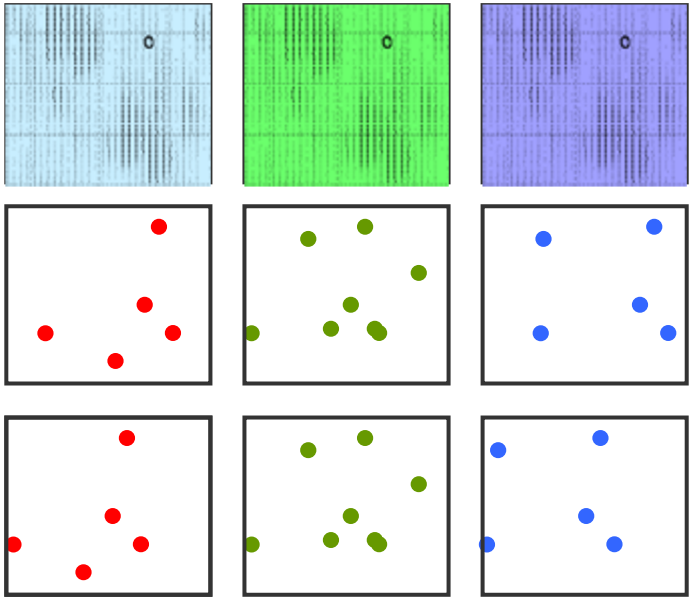
1. Find features in all maps



Oliver
Kohlbacher
University of
Tübingen

Metabolic Profiling

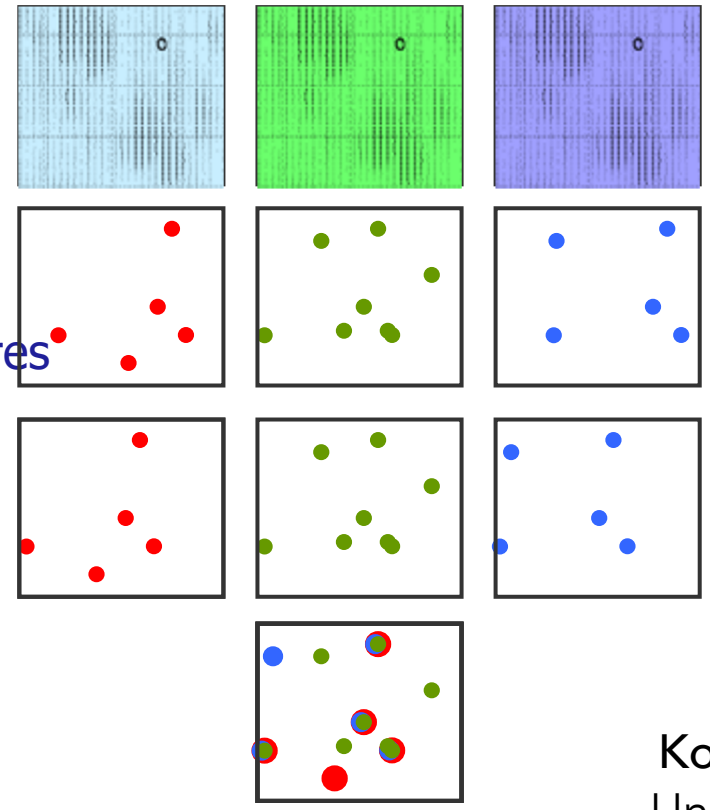
1. **Find** features in all maps
2. **Align** maps



Oliver
Kohlbacher
University of
Tübingen

Metabolic Profiling

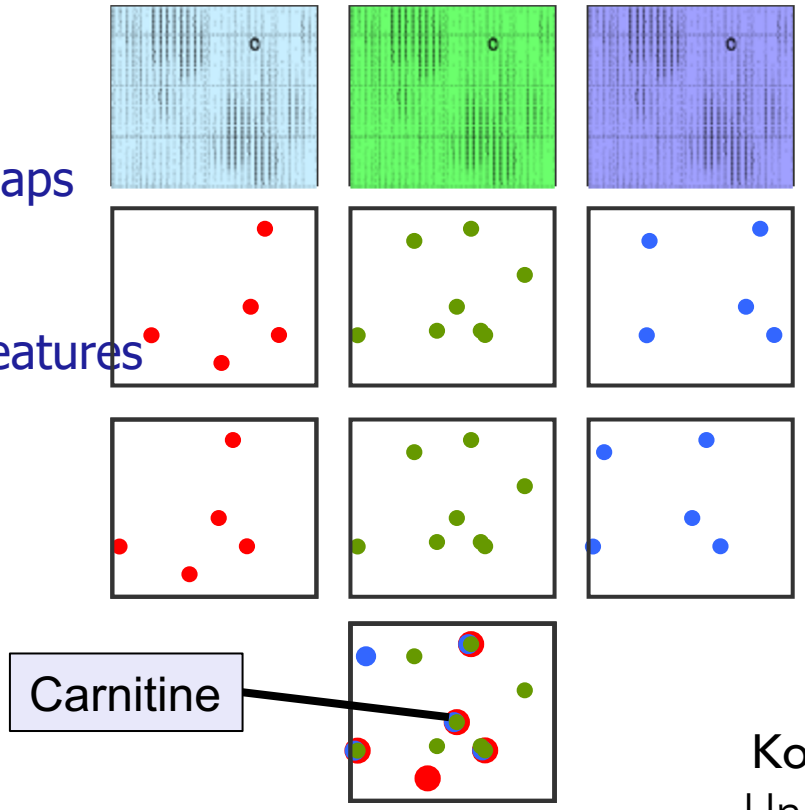
1. **Find** features in all maps
2. **Align** maps
3. **Link** corresponding features



Oliver Kohlbacher
University of Tübingen

Metabolic Profiling

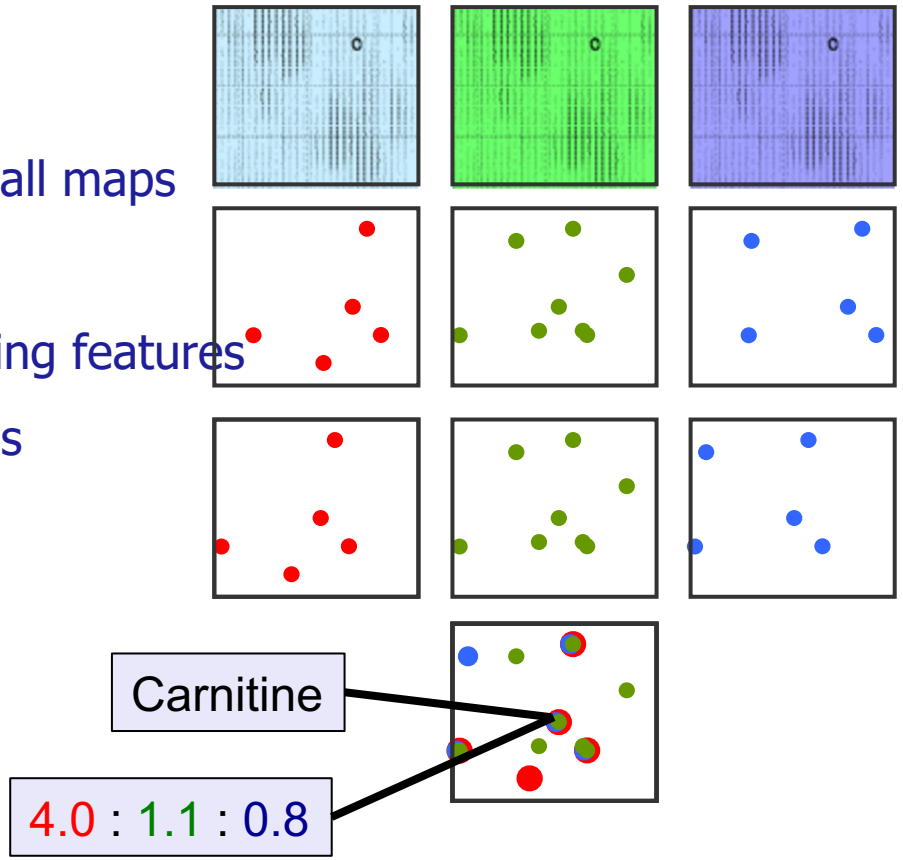
1. **Find** features in all maps
2. **Align** maps
3. **Link** corresponding features
4. **Identify** features

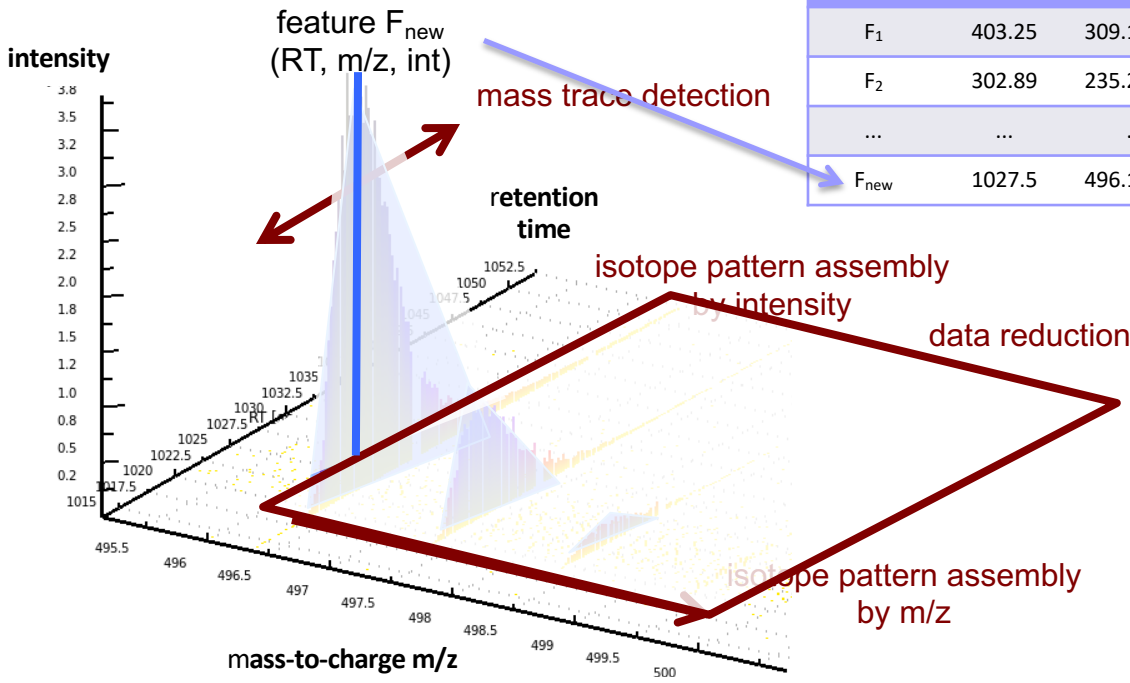


Oliver Kohlbacher
University of Tübingen

Metabolic Profiling

1. **Find** features in all maps
2. **Align** maps
3. **Link** corresponding features
4. **Identify** features
5. **Quantify**



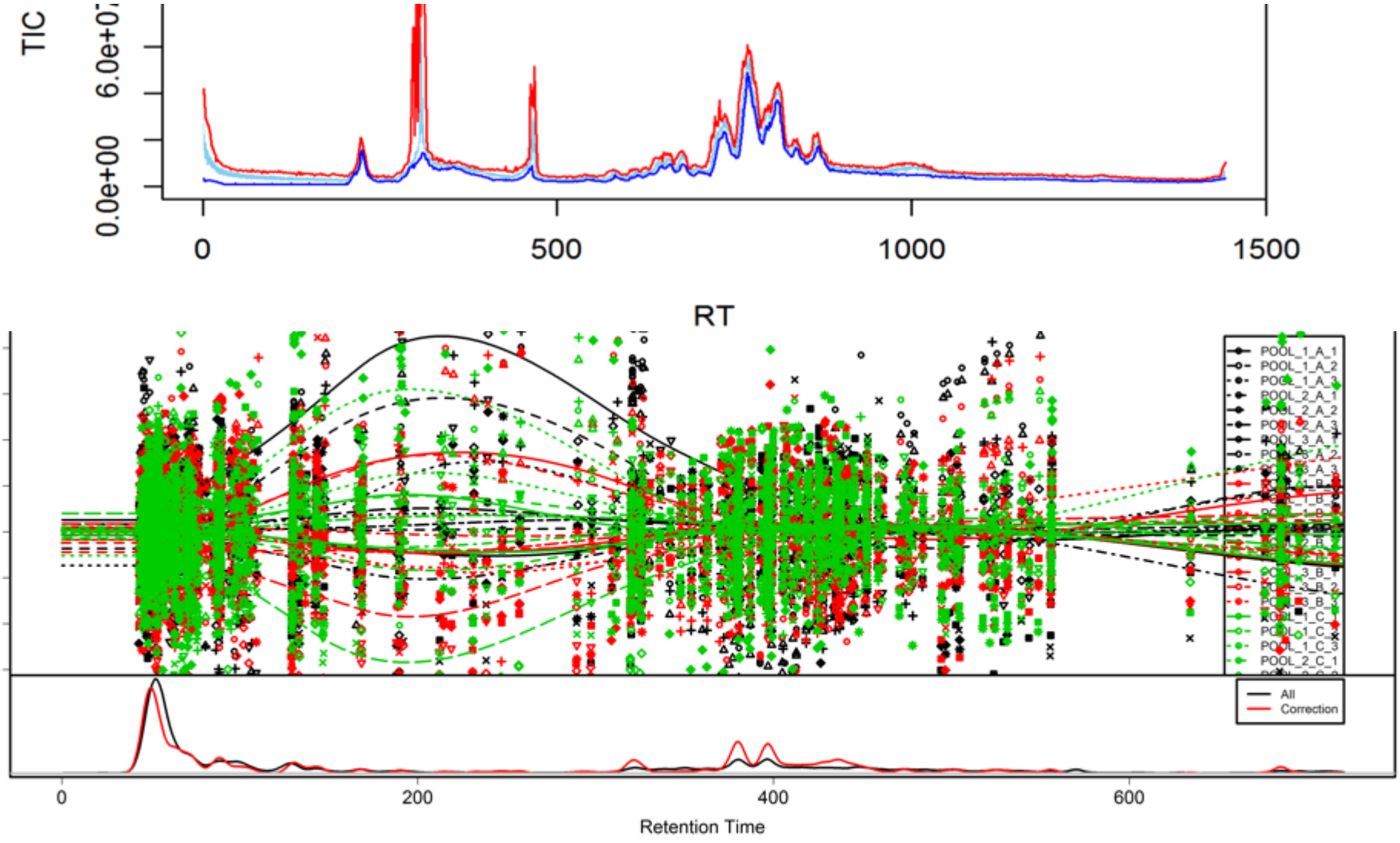


Feature ID	RT	m/z	intensity
F_1	403.25	309.13455	345923.1
F_2	302.89	235.20503	8109.5
...	...	...	...
F_{new}	1027.5	496.11304	45209.8

Oliver Kohlbacher
University of Tübingen

Kenar et al., Mol. Cell. Prot., 2014, 13(1):348-59. doi: 10.1074/mcp.M113.031278

Grouping and retention time correction

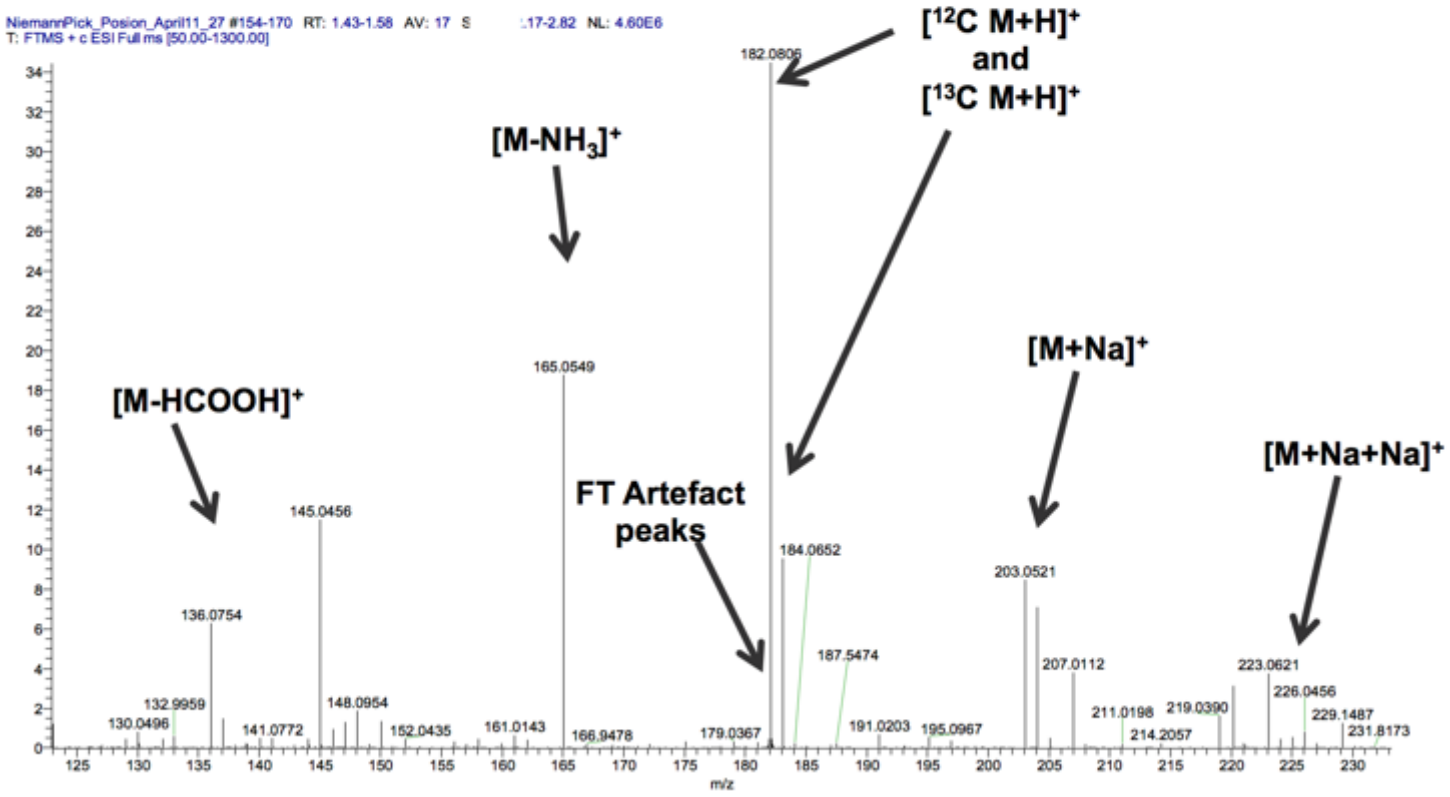


https://cdn.rawgit.com/stanstrup/stanstrup.github.io

World Health Organization

The complexity of data for Tyrosine

17 IONS OF DIFFERENT MASS AND SAME RETENTION TIME CREATED FROM A SINGLE METABOLITE – NEED TO DEFINE THE ION TYPE FOR IDENTIFICATION OR HIGH PROBABILITY OF FALSE POSITIVE IDENTIFICATION



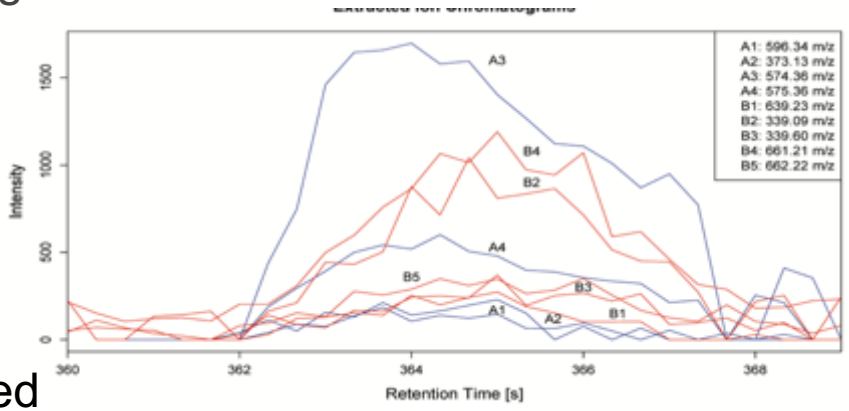
International Agency for Research on Cancer

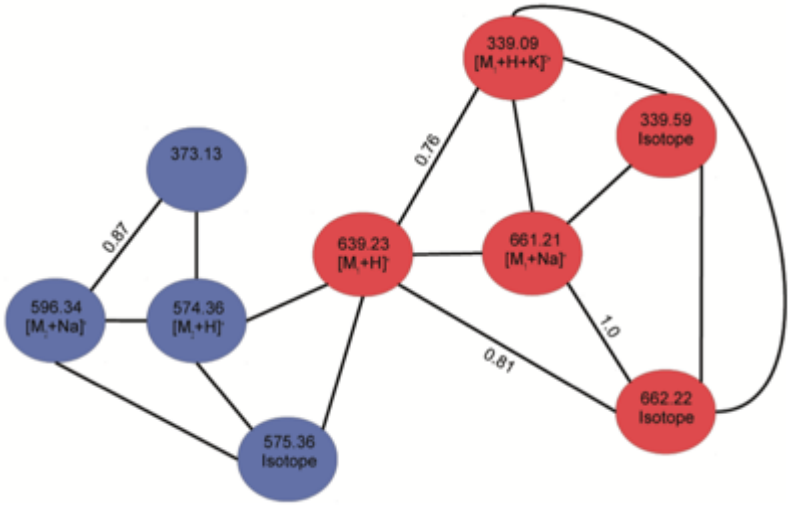
World Health Organization

Slide from Rick Dunn

Camera algorithms

1. **groupFWHM:** Using the most intense features the data is divided into rough retention time groups.
2. **Findisotops:** Look for C12/C13 isotope differences of m/z. Checks if intensity profile matches for [M+]+ to [M+1]+
3. **groupCorr:** Using the extracted ion chromatograms (EIC) for each feature are used to calculate *two types of correlation* between features:
   - Correlation across samples: CAS
   - Correlation within samples : CPSi
4. **Network relationship map** to build:
   - Nodes: features (peak)
   - Weighted Edges: The score (above a threshold)





International Agency for Research on Cancer

World Health Organization

- Creates peak correlation (peak cluster) groups e.g. Red and blue colours

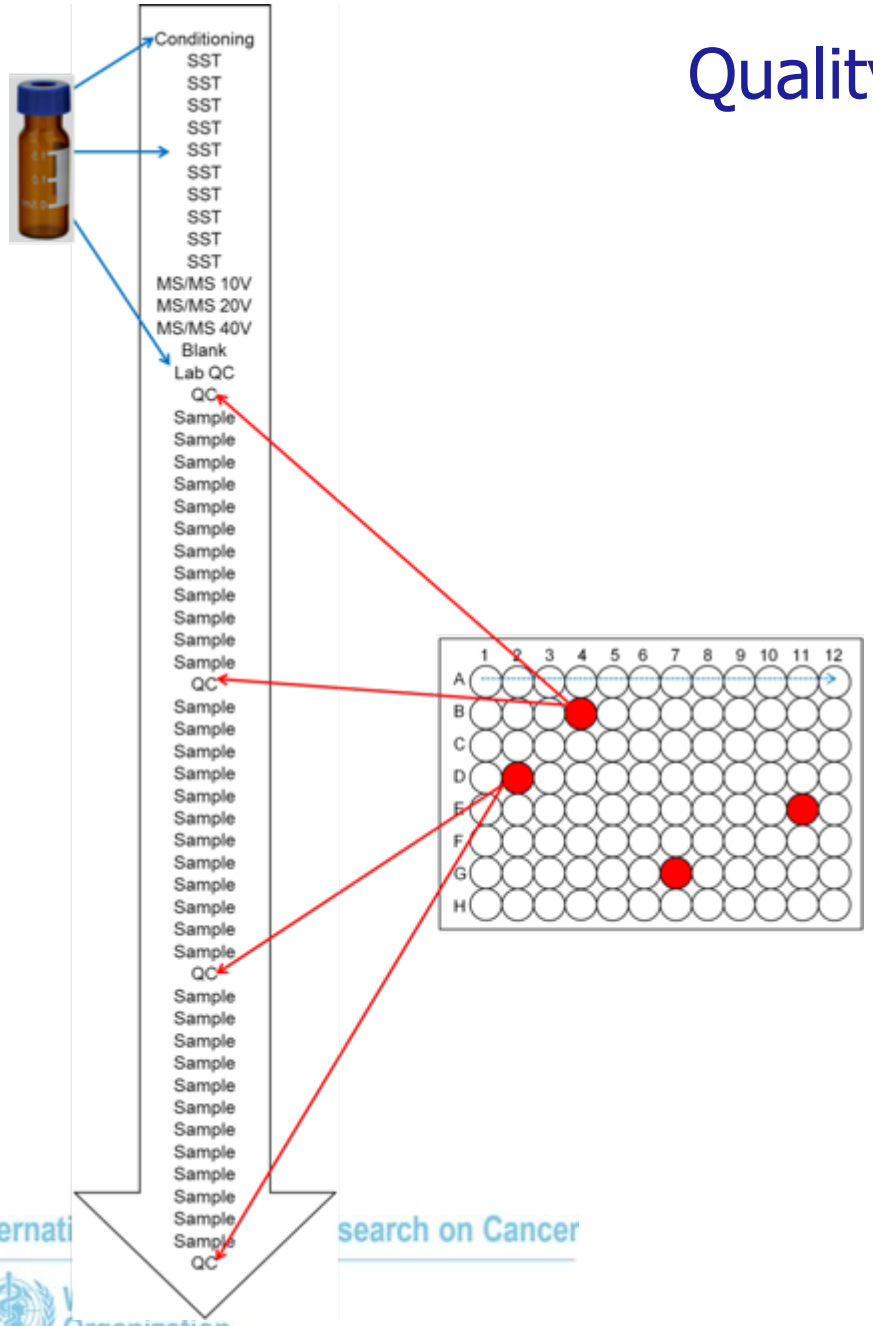
Peak tables and filtering, data clean-up

METABOLOMICS DATA

International Agency for Research on Cancer

World Health Organization

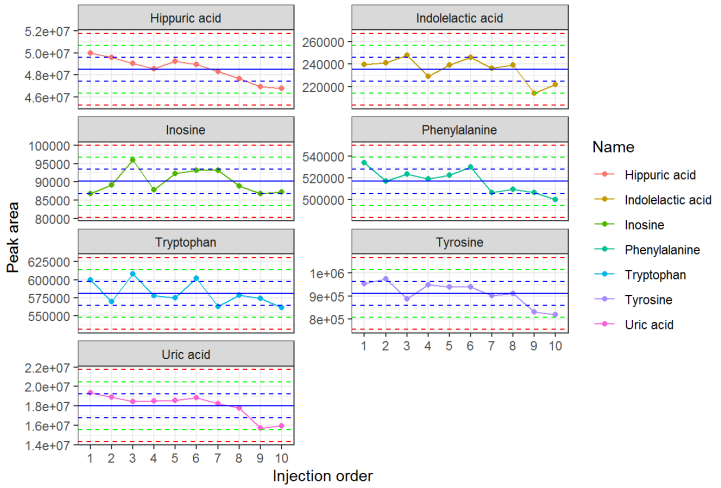
Quality control



QC on preselected metabolites

Metabolite	Mean (peak area)	Peak area RSD (%)
Cholesterol	2358180	6.3
Cortisol	128190	10.9
Creatine	3461812	9.0
Decanoylcarnitine	116241	9.8
Hippuric acid	626165	10.2
Hypoxanthine	1701776	8.7
Indole-3-acetic acid	1923894	8.4
Indolelactic acid	503305	7.7
Indolepropionic acid	399383	10.0
Inosine	36161	12.0
Kynurenine	1278834	11.7
Phenylalanine	27782412	8.9
Retinol	99445	10.3
Tryptophan	30302944	12.1
Tyrosine	2465081	14.4
Valine	43778692	3.3

Westgard rules: target metabolites in QC samples



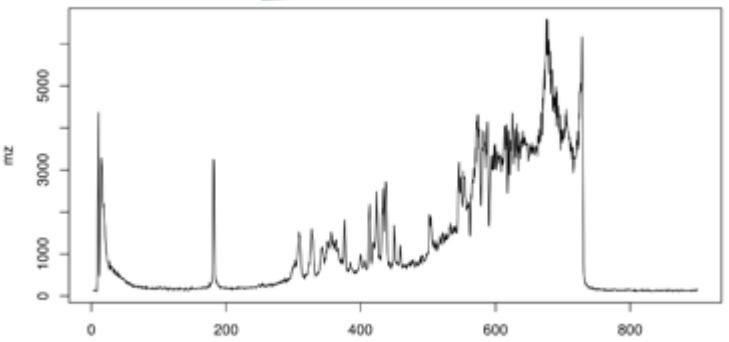
International Agency for Research on Cancer
World Health Organization

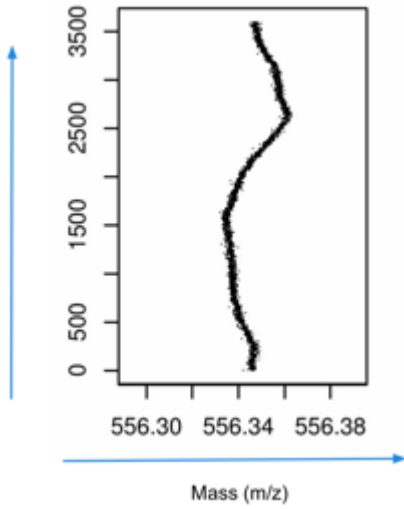
Removing "signals" blanks

The primary purpose of blanks is to trace sources of artificially introduced contamination.

remove remove Keep remove Keep ?

Compound	113.9926@0.578557	147.9439@	151.0079@	166.9856@	171.8773@	172.1088@	181.9796@	188.0133@	188.1032@	188.1398@	190.9378@	191.9349@	195.0879@	202.1189@	213.0086@	215.1506@	222.161@6	224.1395@	228.2079@	234.1606@	236.1762@	243.1815@
Blank	50747	29554	1412	68546	465	10964	5473	2323	3493	757	1454	13028	36459	30508	160	991	4808	4439	493	7459	3398	3195
Blank 3	461026	10780	3527	56538	844	15840	19743	3524	5965	1403	1961	44857	55654	47143	298	395	8513	7390	1605	11213	6125	2626
Blank 2	300011	30032	2210	72509	819	12958	24618	3322	5429	1062	1046	25537	50853	44227	192	192	6577	5879	973	9155	5155	2354
iarcQC	7835102	23927	1456	70391	5509	13068	746802	3958391	9003	11747	2768	45715	44377	43145	553153	508	5935	6133	6947	70040	3173	5552
QC41	6656094	24882	4348	45058	5377	16748	895759	5474397	9151	11092	2124	33895	49488	48248	508024	2659	11446	6859	20292	5890	7867	2818
61-LivCan_127	5317738	23125	1480	71024	5288	10900	824592	1795322	7878	5800	2421	44424	41749	43559	625550	2150	5122	5585	5494	7401	3848	3226
61-LivCan_128	6014473	22687	1920	67389	5161	11005	747943	7303290	7090	13433	2285	39743	43927	44022	583849	2377	4958	6097	3860	6690	3996	3375
61-LivCan_285	5951576	22210	1950	69372	5458	11095	759274	5847130	7677	4261	2050	35039	38996	44741	299742	2237	5541	6634	5495	5999	3542	3682
61-LivCan_286	7670535	26187	1640	76001	6735	10727	698693	2745342	6405	26742	3477	58295	38361	44662	170905	3547	4973	5446	5416	6346	3964	4377
41-LivCan_153	5291769	18510	1688	73602	6233	10992	787727	6209722	6912	7944	2809	33948	37417	43242	892429	6532	5273	5920	5895	5126	3910	3733
41-LivCan_154	7376605	20030	1800	69293	4923	11228	810178	5745146	7255	8495	2739	39595	38582	46691	479766	5577	5153	5358	4702	5132	4040	3490
41-LivCan_267	5683423	21518	2059	69933	5583	11233	799875	3953747	7262	4129	2233	39406	40907	42595	144910	5424	561					
41-LivCan_268	6169464	20251	2123	69637	5676	10894	793765	2717762	7498	6333	2944	35970	40180	45635	394338	7129	534					
61-LivCan_115	5174324	25655	1974	65465	7371	10686	797960	10239480	8018	14090	4050	46741	42471	45228	377759	2363	482					
61-LivCan_116	5967438	21184	1819	69125	5096	10743	793788	6169026	10060	35472	2374	36452	38537	44304	692925	2256	465					
QC41	8524102	22853	2136	68664	6525	10290	850190	5148956	8381	11131	2360	27451	43934	40927	470808	2935	519					
21-LivCan_299	5747001	21371	1856	61384	4874	10445	805036	3901179	7438	8041	2659	37263	44918	44146	258328	1131	575					
21-LivCan_300	8000284	17898	2139	69604	6817	11194	803831	2556924	7116	8063	2730	40192	41648	42824	308635	864	544					
61-LivCan_363	5471551	16411	2261	69993	5125	11213	771251	4096973	6667	5595	2720	36057	38129	44469	333414	2100	512					
61-LivCan_364	8272500	18046	1980	65601	7511	10957	754949	6258103	6836	12412	2386	30613	38967	43326	222530	1985	522					
22-LivCan_85	5416749	20314	2211	65178	7421	11081	792472	240338	7266	4399	2189	37007	38425	42069	151603	2112	531					
22-LivCan_86	7626831	26384	2161	69751	5284	10091	763346	4175476	7026	4987	2970	44838	40249	43394	707391	1706	516					
25-LivCan_309	5556264	15113	2265	67062	6856	12102	799966	3469769	9575	3408	2374	36101	35924	45377	510291	2728	526					
25-LivCan_310	7974036	21822	1949	65129	4576	13594	796971	1928580	8305	3557	2160	38242	34928	46392	441340	3155	594					
72-LivCan_305	5522104	18230	1906	68239	5808	10310	786301	1421086	6314	7017	2047	35015	37939	46016	563617	4443	520					
72-LivCan_306	12703038	18226	2281	65685	4396	11262	811692	83133	7330	9013	2370	25679	39841	46120	472587	1988	554					
41-LivCan_263	8111869	18572	1445	66156	5716	10581	790852	16321126	8670	8324	2590	31625	40002	47278	505102	3898	523					
41-LivCan_264	5188782	20548	1166	66497	5479	10773	791031	8766933	7921	7032	2572	37160	43488	41627	622548	3238	561					
QC41	8652256	21088	2780	66314	7715	10542	851495	5432191	8172	10588	2402	38622	46389	41625	492936	3151	563					
42-LivCan_13	7949	29872	1962	65466	4601	10930	758754	4554789	7881	11871	2267	33204	44192	47555	831662	1304	488					
								4435128	7880	7165	1952	23310	41866	45472	211016	1104	517					
								5103886	6854	11399	1696	26586	38270	43359	353765	847	606					
								11355299	6616	9724	2530	33255	36287	46594	245268	833	569					
								215596	35396	6183	1840	31761	39576	45809	25102	831	566					
								3132387	37202	4139	2221	35868	39307	45550	98013	727	565					
								15299321	7755	8223	2114	35351	37262	45451	301054	2076	534					
								264641	7451	4582	1997	19769	38256	46611	76608	2383	605					
								30885852	8366	6461	2469	36785	39877	44513	1672841	2260	561					

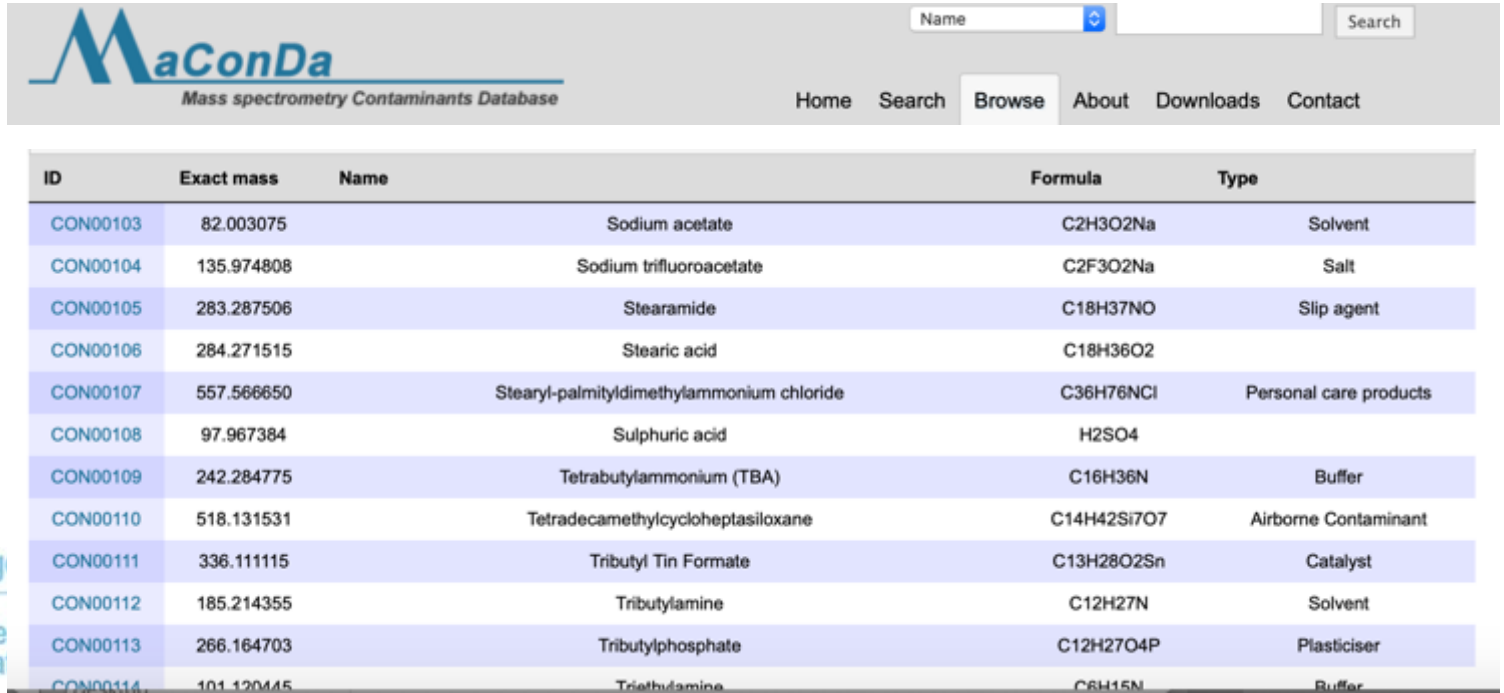




MaConDA

Method Blank: A blank prepared to represent the matrix as closely as possible. The method blank is prepared/extracted/digested and analysed exactly like the field samples.
Purpose: Assess contamination introduced during sample preparation

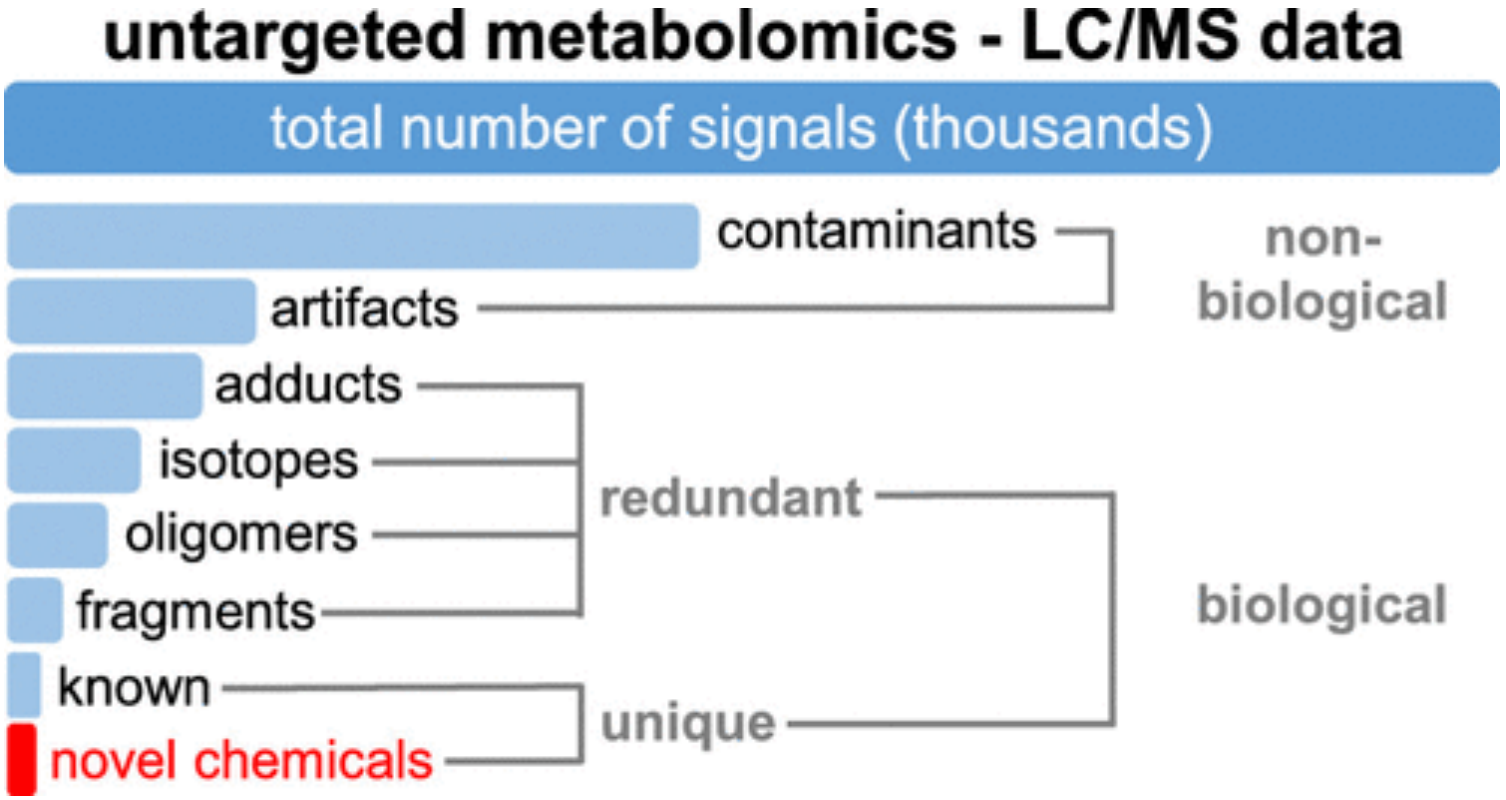
Instrument Blank: A blank analysed with field samples.
Purpose: Assess the presence or absence of instrument contamination.
Frequency: Defined by the analytical method or at the analyst's discretion (e.g., after high concentration samples).



ID	Exact mass	Name	Formula	Type
CON00103	82.003075	Sodium acetate	C2H3O2Na	Solvent
CON00104	135.974808	Sodium trifluoroacetate	C2F3O2Na	Salt
CON00105	283.287506	Stearamide	C18H37NO	Slip agent
CON00106	284.271515	Stearic acid	C18H36O2	
CON00107	557.566650	Stearyl-palmityldimethylammonium chloride	C36H76NCl	Personal care products
CON00108	97.967384	Sulphuric acid	H2SO4	
CON00109	242.284775	Tetrabutylammonium (TBA)	C16H36N	Buffer
CON00110	518.131531	Tetradecamethylcycloheptasiloxane	C14H42Si7O7	Airborne Contaminant
CON00111	336.111115	Tributyl Tin Formate	C13H28O2Sn	Catalyst
CON00112	185.214355	Tributylamine	C12H27N	Solvent
CON00113	266.164703	Tributylphosphate	C12H27O4P	Plasticiser
CON00114	101.120445	Triethylamine	C6H15N	Buffer

International Ag
World He
Organiza

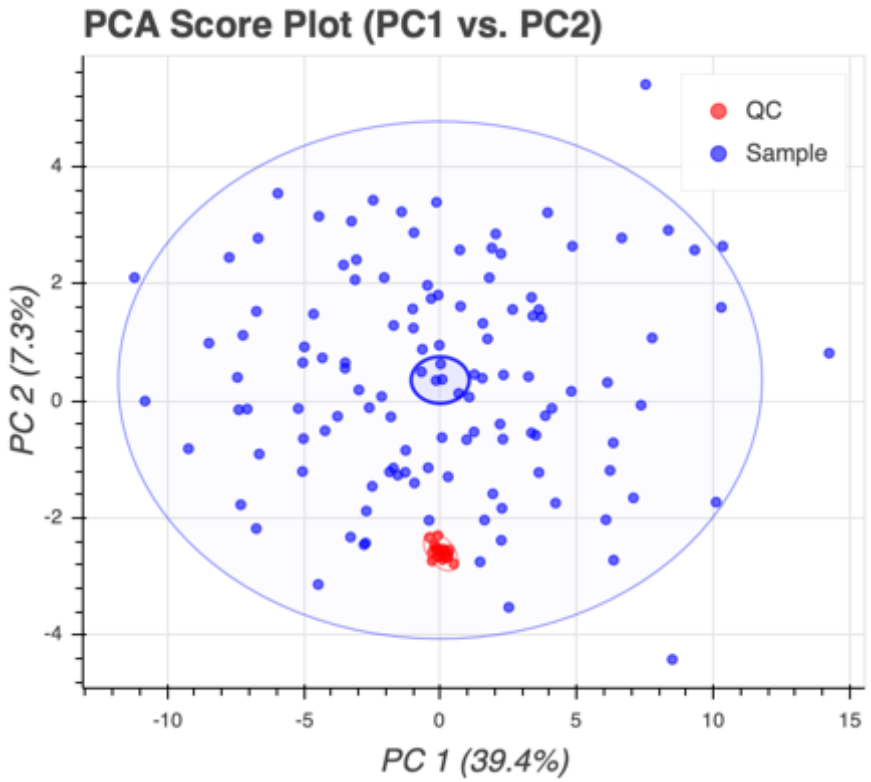
Metabolomics Data reduction

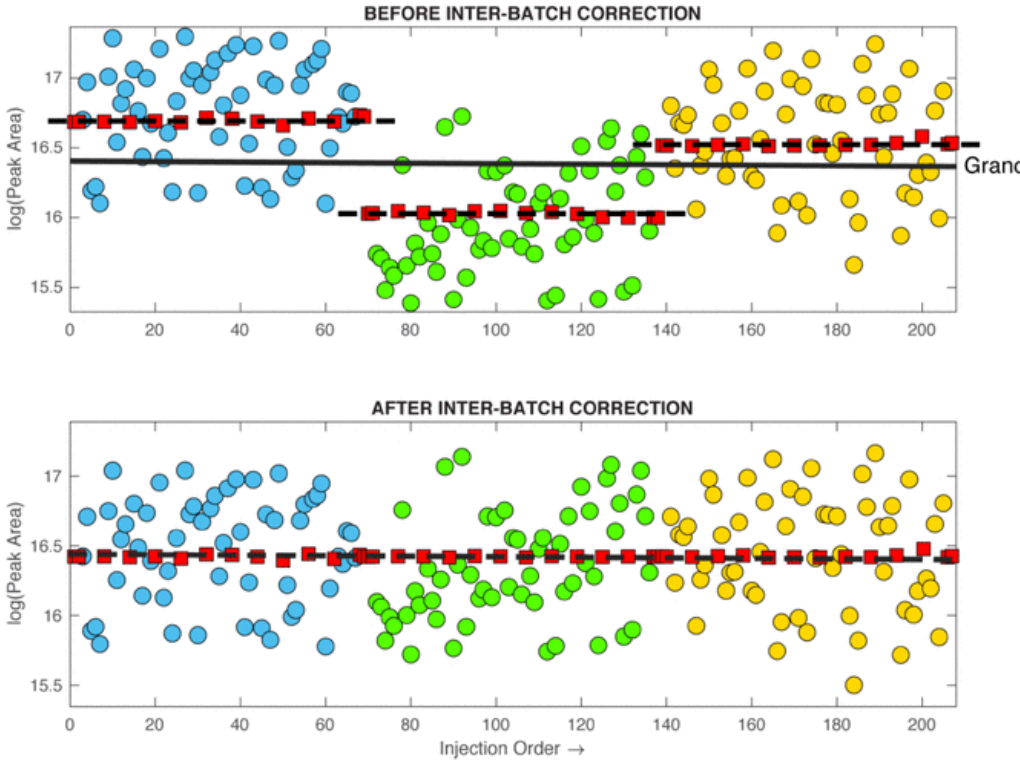


https://doi.org/10.1021/jacs.9b13198

https://yufree.github.io/pmd/

QC and Batch correction





https://doi.org/10.1007/s11306-018-1367-3

International Agency for Research on Cancer

World Health Organization

Metabolite identification

METABOLOMICS DATA

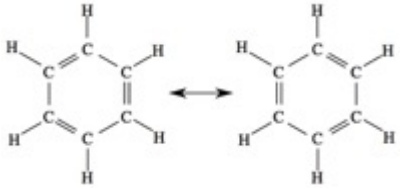
International Agency for Research on Cancer

World Health Organization

Metabolome - Dark Matter

Dynamic Range and Instrument Sensitivity

Heterogeneity



multiple methods for sample prep and analysis

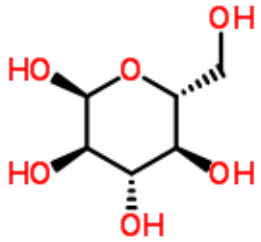
Chemical Lability

Keep samples cold, minimum prep

Chemical Instability

Ionization Techniques and choice of MS

Isomers



Glucose (1 formula, 1 accurate mass)
385 structures

Adducts, fragments

Mass spectrometers are machines for performing chemistry – Graham Cooks!

Database search

https://hmdb.ca/

Jump To Section: Identification | Taxonomy | Ontology | Physical properties | Spectra | Biological properties | Concentrations | Links | References | enzymes (2) | Show 2 proteins | XML

GC-MS	GC-MS Spectrum - GC-EI-TOF (Pegasus III TOF-MS system, Leco; GC 6890, Agilent Technologies) (Non-derivatized)	splash10-0002-0940000000-3ab0a8161f2ec8f04452	Spectrum
GC-MS	GC-MS Spectrum - GC-EI-TOF (Pegasus III TOF-MS system, Leco; GC 6890, Agilent Technologies) (3 TMS)	splash10-00dj-9630000000-1dc322333cb48ef39588	Spectrum
GC-MS	GC-MS Spectrum - GC-MS (3 TMS)	splash10-004i-3591000000-	Spectrum

Show more...

Biological Properties

Cellular Locations

- Cytoplasm

Biospecimen Locations

- Blood
- Breast Milk
- Cerebrospinal Fluid (CSF)
- Saliva
- Sweat
- Urine

Tissue Locations

- Prostate

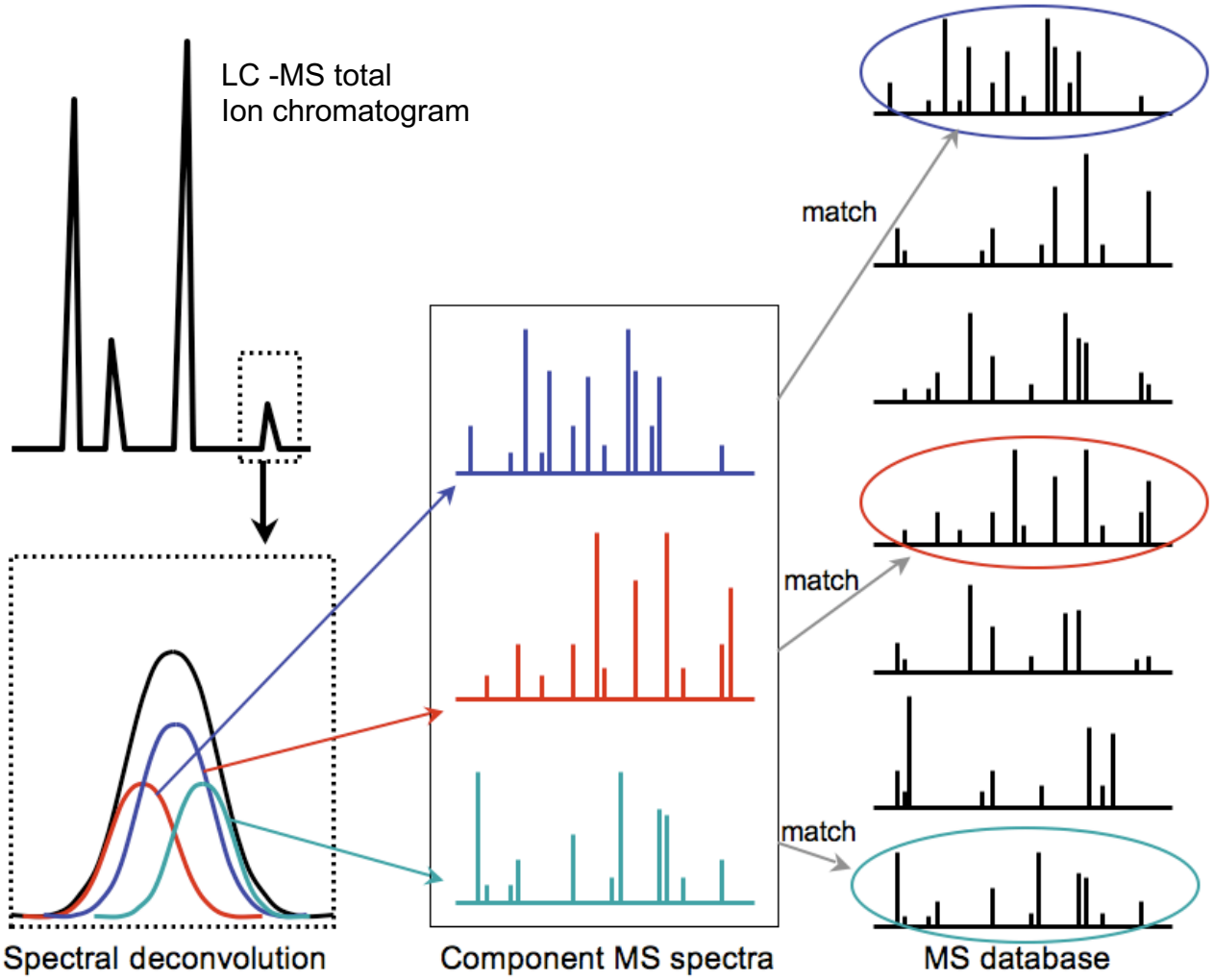
Pathways

Name	SMPDB/Pathwhiz	KEGG
Citric Acid Cycle		
Congenital lactic acidosis		Not Available
Fumarase deficiency		Not Available

InChI Key GTZCVFVGUGFEME-IWQZZHSRSA-N

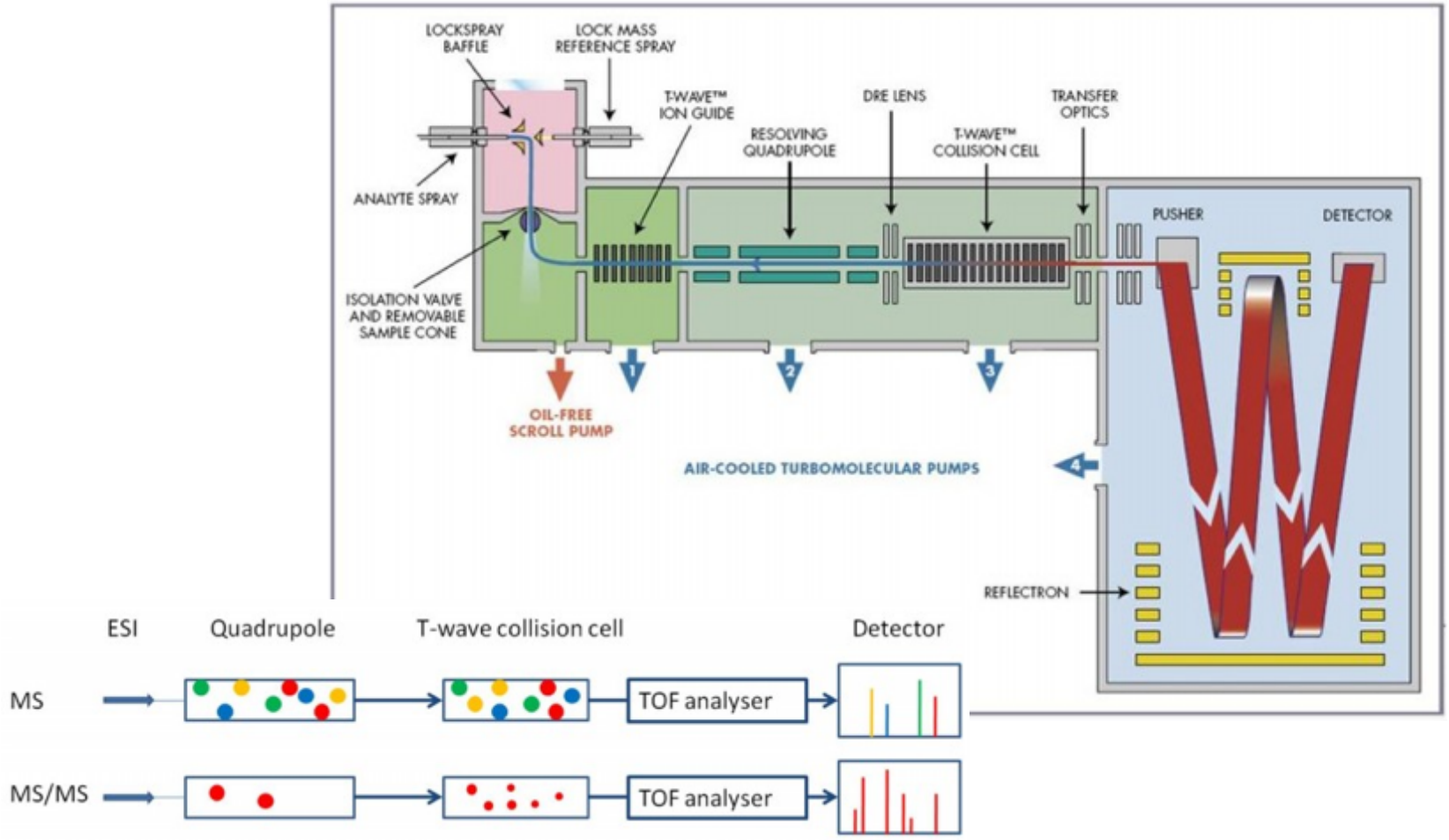
Organization

Metabolite ID by LC-MS



International Age
World Health Organization

Metabolite ID using MS/MS



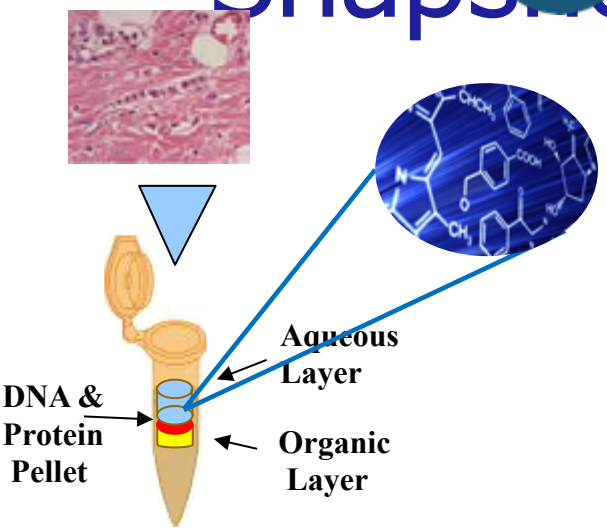
Data tables and statical analysis

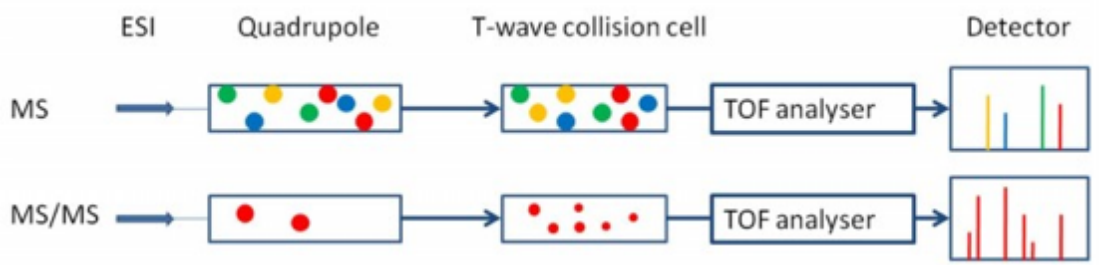
METABOLOMICS ANALYSIS

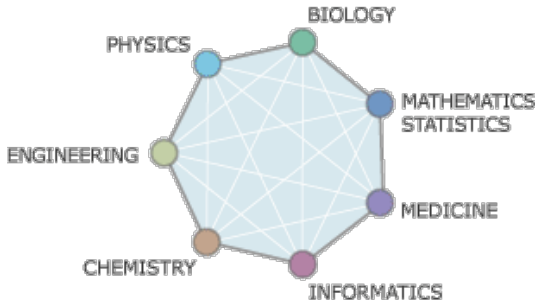
International Agency for Research on Cancer

World Health Organization

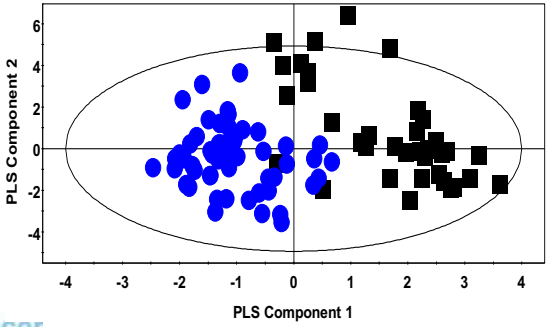
Snapshot metabolomics

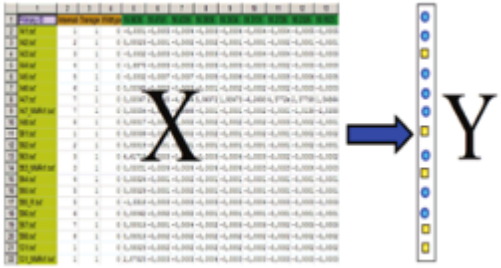






Oltvai-Barabasi, Science, Oct 02



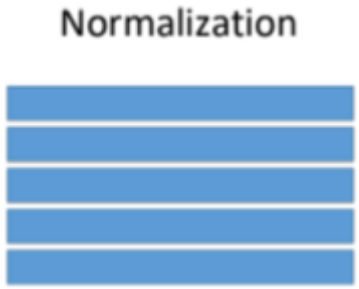


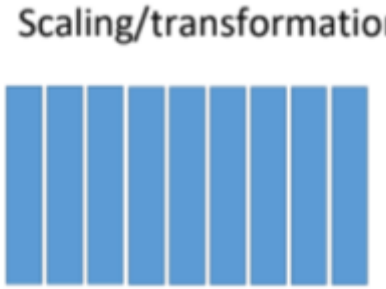
International Agency for Research on Cancer

World Health Organization

Preparing the data for statistical analysis

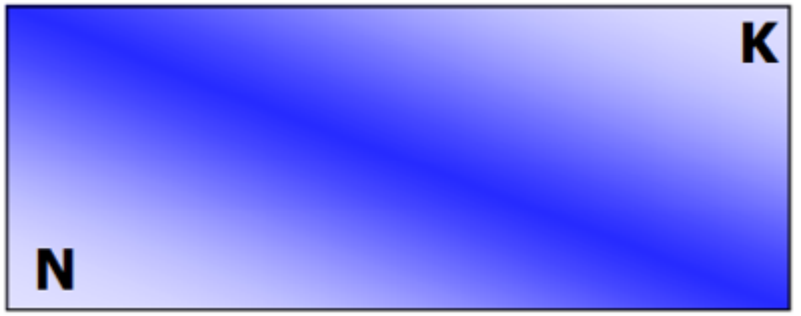
- **Data Normalization**: Making **each observation comparable** to others by **removing effects due to systematic biases**, which could originate from e.g., overall concentration differences or other systematic effects.
  - Total area normalization (TAN)
  - Total sum normalization (TSN)
  - Normalization to standard
  - Normalization to the major peak
  - Probabilistic Quotient Normalization (PQN)
  - Quantile normalization (QN)





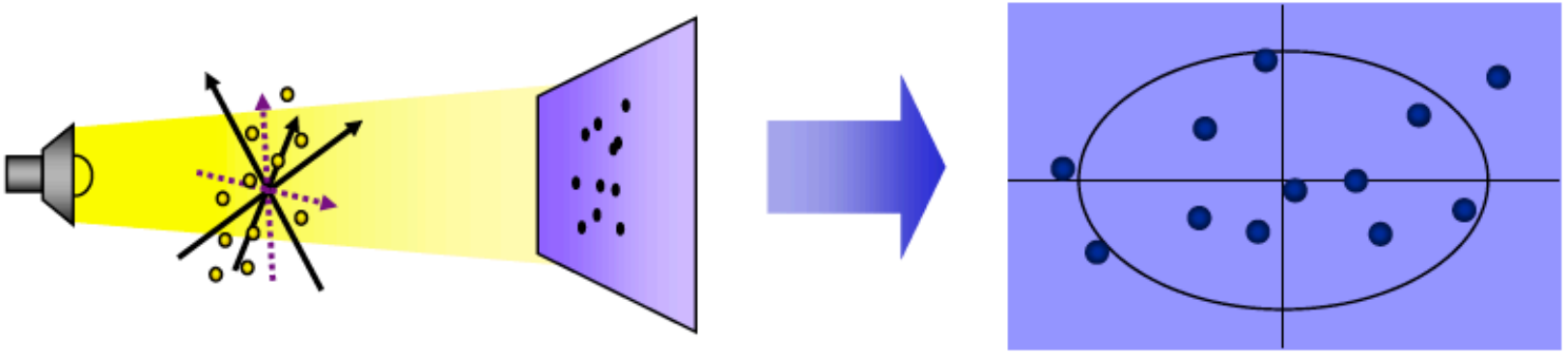
- **Data Centring**
- **Data Scaling**
  - Scaling based on data dispersion
  - Autoscaling, Pareto scaling, range scaling, vast scaling
- Scaling based on average value
  - Level scaling
- **Data Transformations**

PCA and MVDA

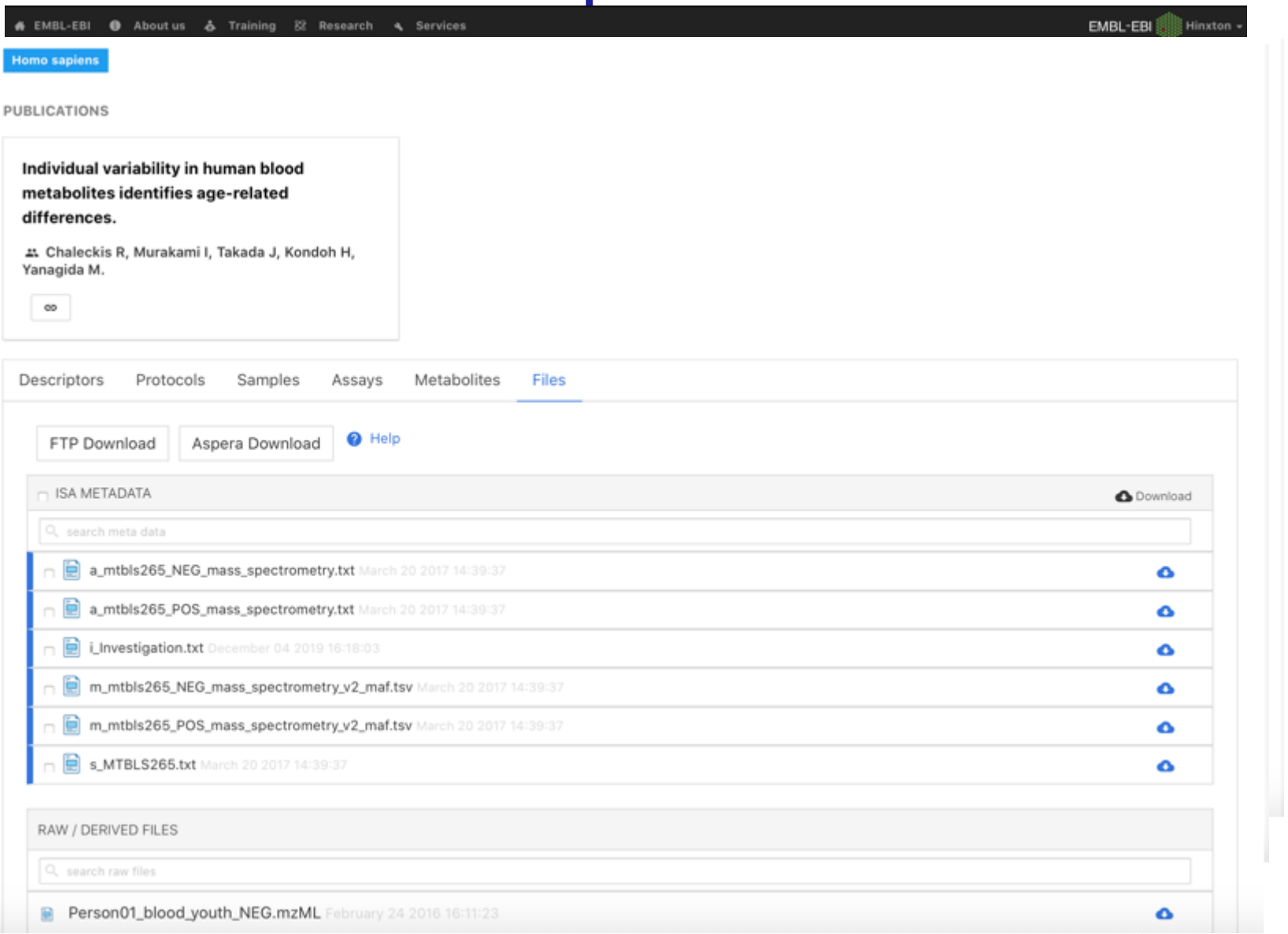


Multivariate analysis by Projection

- Looks at ALL the variables together
- Avoids loss of information
- Finds underlying trends = "latent variables"
- More stable models



Metabolomics experimental database



2012

Internatic

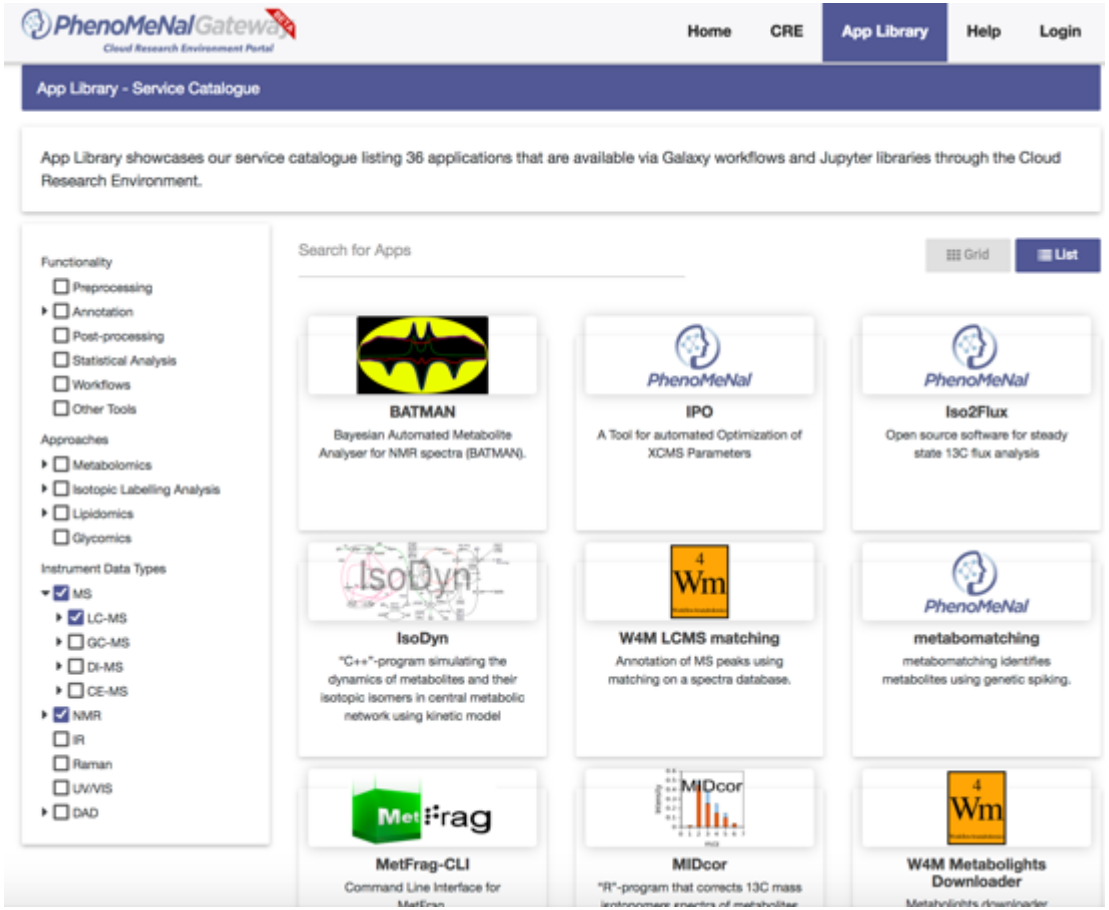
World Health Organization

https://www.ebi.ac.uk/metabolights/

isatab

Risa

Metabolomics tools; PhenoMeNal



https://portal.phenomenal-h2020.eu/

International Agency for Research on Cancer

World Health Organization

Acknowledgments

Biomarkers Group
Augustin Scalbert
Roland WEDEKIND
Vanessa Neveu
Genevieve Nicolas
Laure DOSSUS
Manon CAIRAT
Pekka Keski-Rahkonen
Nivonirina Robinot
David Achaintre
Sabina Rinaldi
Audrey Gicquiau
Mathilde His
Anne-Sophie NAVIONIS
Reza Salek
Adam Amara
Parisa Shahnazari
Pedro Ruas
Vanessa Neveu

Collaboration
Steffen Neumann
Fabien Jourdan
Francisco Couto
Justin van der Hooft
Kati Hanhineva
David Wishart

rch on Cancer

World Health Organization

EPIC PIs