

Metabolomics

Kim Kultima, Assoc. Prof.
Uppsala University
Uppsala University Hospital
Karolinska Institutet



Uppsala University Hospital, Sweden

CARAMBA

CARAMBA- Clinical Analysis & Research Applying Mass spectrometry & Bioinformatics at Akademiska

- Joint venture Uppsala University and Uppsala University Hospital
- The facility is certified (ISO 15189) enabling us to deliver results for clinical care

Clinical metabolomics and proteomics

- ***Multiple sclerosis, Huntington's disease, Alzheimer's disease and chronic pain diseases***



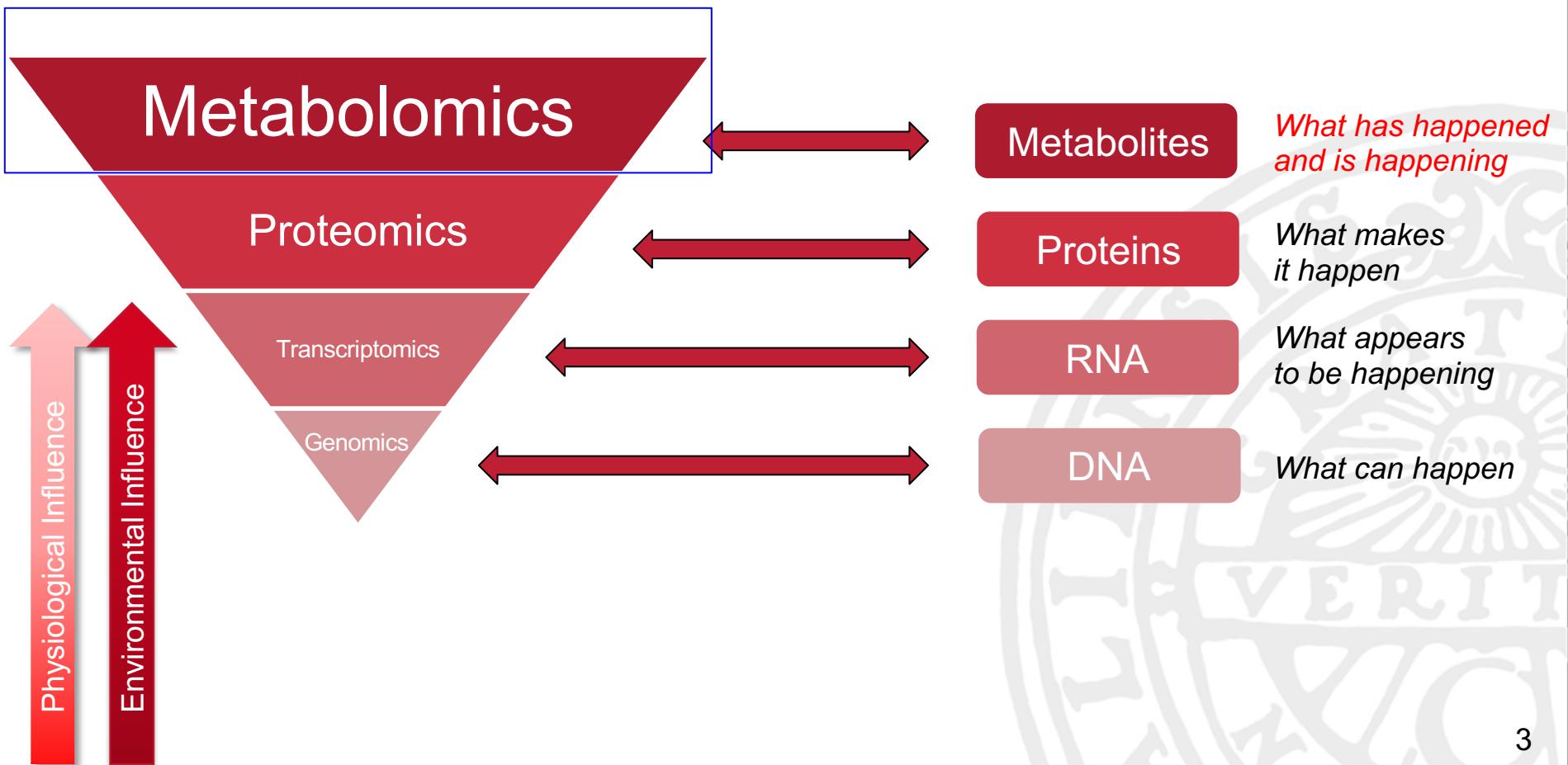
Uppsala University Hospital, Sweden



Shared laboratory for clinical- and research use only using HR-MS

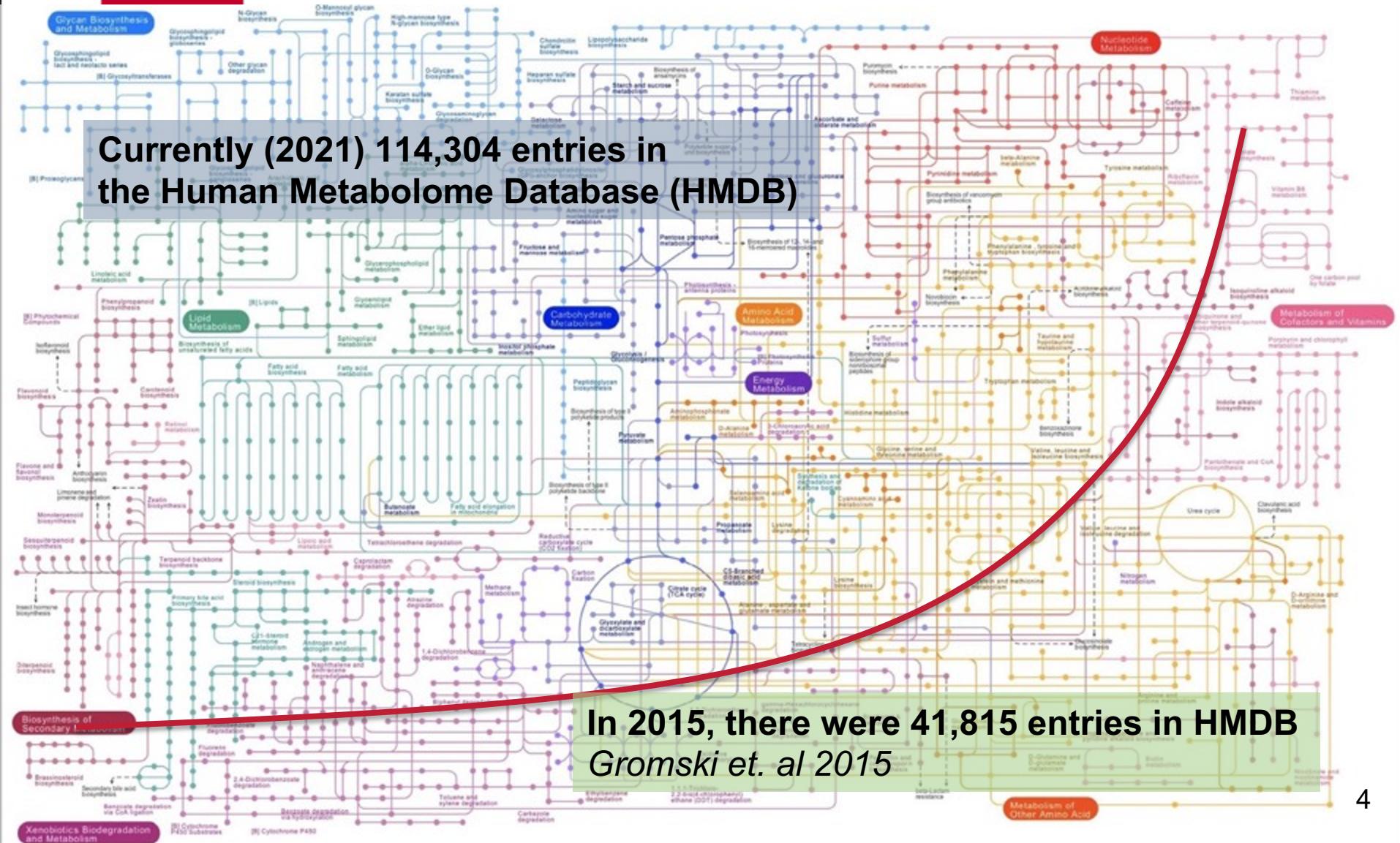
Metabolomics

The study of small molecules, typically <1500 Da in size





Metabolomics

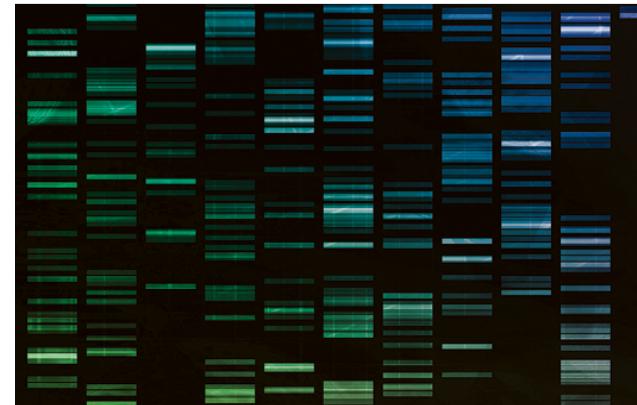


Big Data in Metabolomics - Today

Big data, big picture: Metabolomics meets systems biology

Metabolomics—the study of the collection of an organism's metabolites—provides a molecular measurement of phenotype, or the characteristics resulting from the genotype's interaction with the environment. Using a range of analytical tools to scale the mountains of data collected, including molecular detection and bioinformatics, scientists use metabolomics to understand systems biology, which is the complete computational analysis and modeling of an organism and its well-being. **By Mike May**

Despite all of the advances in storing and analyzing data, scientists are still confronting significant obstacles in studying the metabolome. “One bottleneck in nontargeted workflows is the identification of unknown compounds,” says Aiko Barsch, market manager for metabolomics at **Bruker Daltonics**, based in Bremen, Germany. “This is where MS and NMR [mass spectrometry and nuclear magnetic resonance] both have advantages.”



“Metabolomics can't really function on its own without the genome sequence.” — Jonas Korlach, chief scientific officer at Pacific Biosciences

Big Data in Metabolomics - In health

Predicting human health from biofluid-based metabolomics using machine learning

Ethan D. Evans¹, Claire Duvallet^{1,3}, Nathaniel D. Chu¹, Michael K. Oberst²,
Michael A. Murphy^{1,2}, Isaac Rockafellow^{1,4}, David Sontag^{1,2*} & Eric J. Alm^{1,2*}

Biofluid-based metabolomics has the potential to provide highly accurate, minimally invasive diagnostics. Metabolomics studies using mass spectrometry typically reduce the high-dimensional data to only a small number of statistically significant features, that are often chemically identified—where each feature corresponds to a mass-to-charge ratio, retention time, and intensity. This practice may remove a substantial amount of predictive signal. To test the utility of the complete feature set, we train machine learning models for health state-prediction in 35 human metabolomics studies, representing 148 individual data sets. Models trained with all features outperform those using only significant features and frequently provide high predictive performance across nine health state categories, despite disparate experimental and disease contexts. Using only non-significant features it is still often possible to train models and achieve high predictive performance, suggesting useful predictive signal. This work highlights the potential for health state diagnostics using all metabolomics features with data-driven analysis.

Rise of Deep Learning for Genomic, Proteomic, and Metabolomic Data Integration in Precision Medicine

Dmitry Grapov¹, Johannes Fahrmann^{2,*}, Kwanjeera Wanichthanarak^{3,4,*} and Sakda Khoomrung^{3,4}

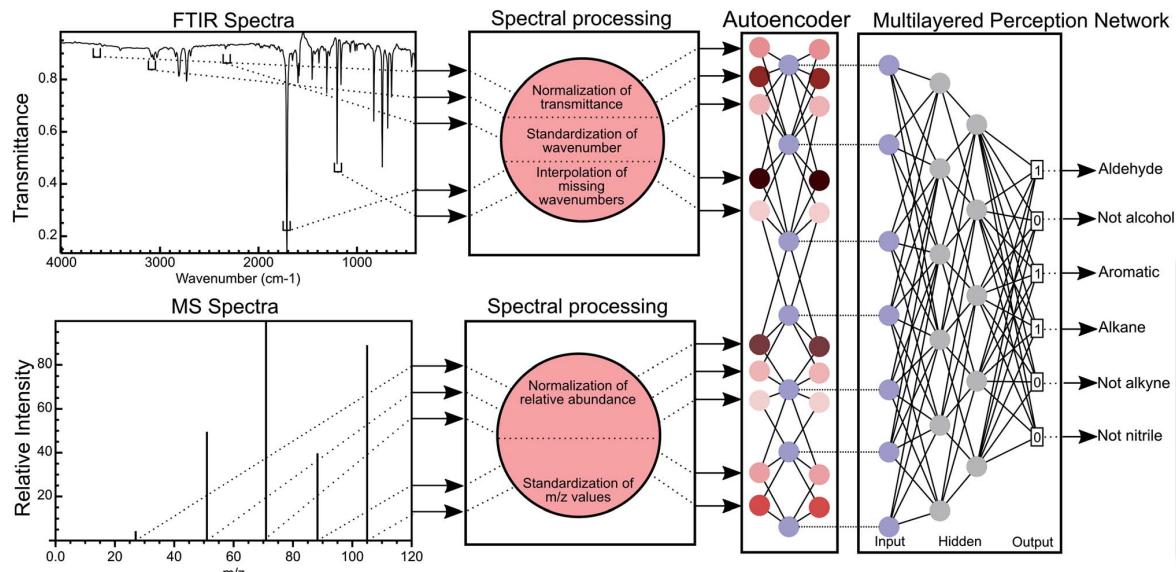
Big Data in Metabolomics - Out reach

Spectral deep learning for prediction and prospective validation of functional groups†

Jonathan A. Fine, ‡^a Anand A. Rajasekar, ‡^b Krupal P. Jethava ^a
and Gaurav Chopra

Deep learning to generate *in silico* chemical property libraries and candidate molecules for small molecule identification in complex samples

Sean M. Colby, Jamie R. Nufiez, Nathan O. Hodas, Courtney D. Corley, Ryan R. Renslow*
Pacific Northwest National Laboratory, Richland, WA, USA.
* ryan.renslow@pnnl.gov



Analytical platforms for metabolomics

- Metabolites have a vast range of chemical structures, properties and concentrations
- No single platform provides complete comprehensive coverage
- No single extraction or analysis method works for all metabolites
- Selection of the platform is always a compromise between sensitivity, speed and chemical selectivity and coverage

Common techniques used in metabolite profiling studies:

- NMR
- MS

Hyphenated techniques

- GC-MS
- CE-MS
- LC-MS

Big Data in metabolomics- Challenges

- The rapid progress in field has resulted in a mosaic of independent, and sometimes incompatible, analysis methods that are difficult to connect into a useful and complete data analysis solution
- Configuring necessary software tools and chaining them together into a complete re-runnable analysis is challenging

NMR vs MS-based Methods

NMR (Nuclear Magnetic Resonance)

- Untargeted approach



MS (Mass spectrometry)

- Targeted or untargeted



Untargeted (global) approach

- Measures as many metabolites as possible from a range of biological samples



Targeted (specific) approach

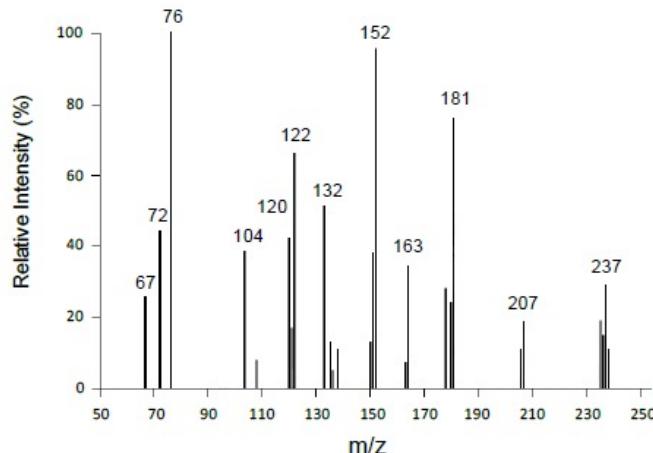
- To measure sets of metabolites when you have a specific biochemical question you want to answer

Comparison of NMR and MS

NMR Spectroscopy	Mass Spectrometry
Less sensitive for metabolite detection	More sensitive for metabolite detection
Non-destructive	Destructive
Fast	Moderate (LC-MS/GC-MS) to fast (direct infusion)
Requires little sample handling and preparation	Sample preparation/derivatization (GC-MS)
Quantification easy	Can be quantitative
Identification	More difficult to identify new metabolites
Robust	Less robust
Large sample volume (0.1 – 0.5 mL)	Smaller sample volume (10 – 100 μ L)
Fewer metabolites detected	Broad metabolite coverage
Mostly nonselective	Selective and nonselective
High initial instrument cost	Lower initial instrument cost

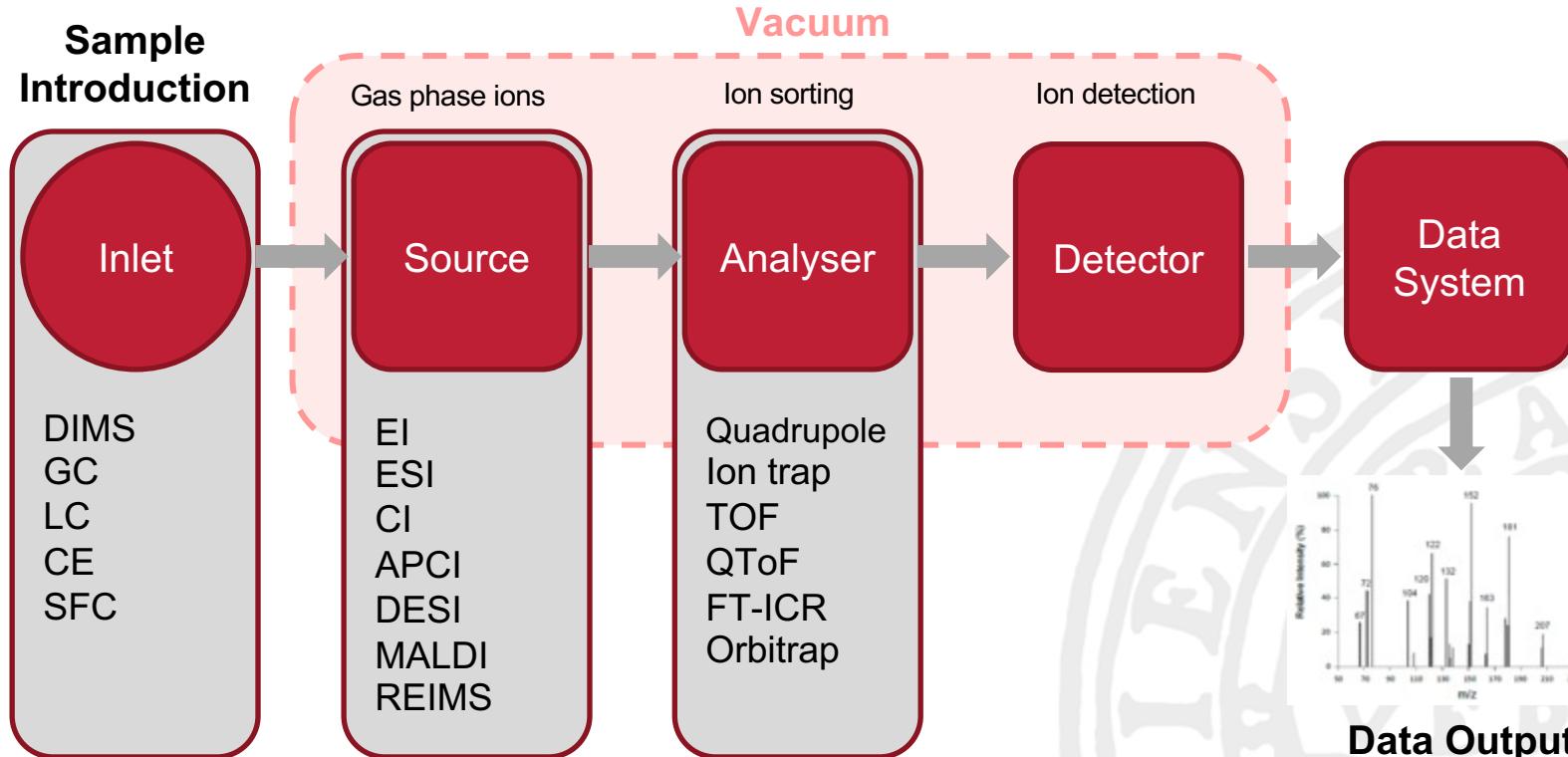
Mass spectrometry

- A tool for ionising molecules
- Sorts the ions based on their mass-to-charge (m/z) ratio
- In simple terms it measures the masses and abundance within a sample
- Qualitative/quantitative detection of molecular ions



Typical mass spectrum

Mass spectrometry



What is Electrospray Ionisation?

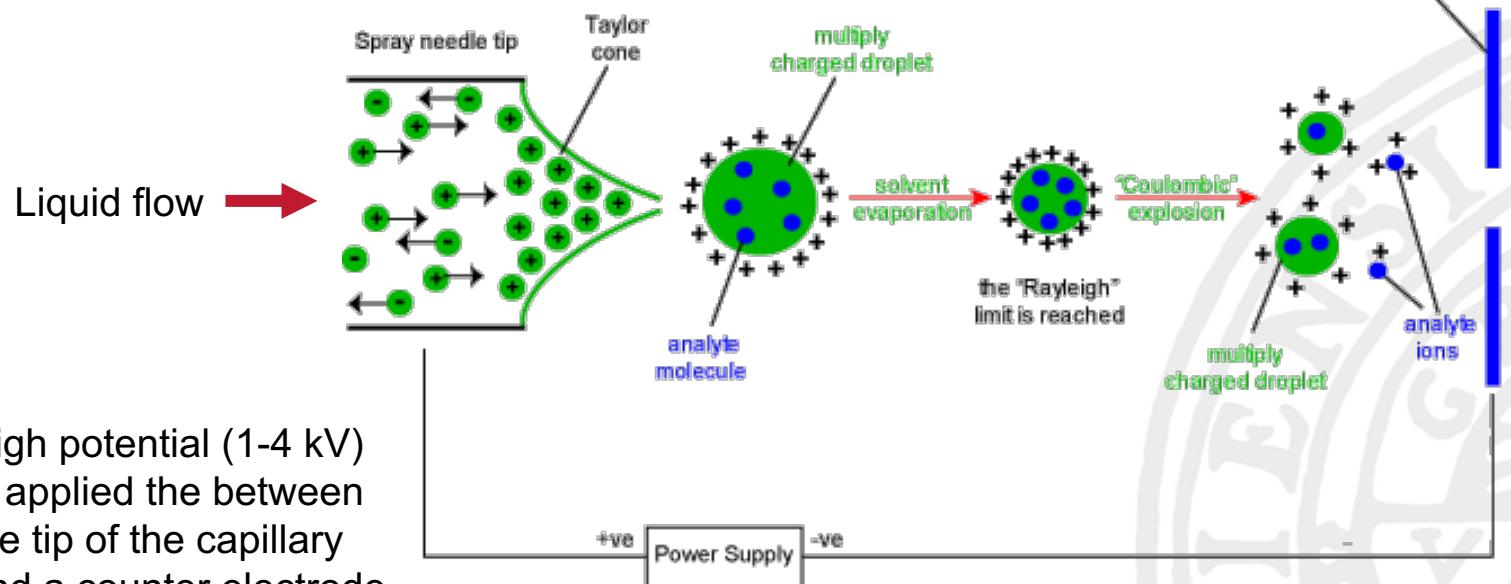
- A technique used to produce ions
- A techniques used to transfer ions from the solution phase into the gas phase
- ESI is a soft ionisation process
- Generates a molecular ion
 - $[M+H]^+$ or $[M-H]^-$
 - In metabolomics it can also be $[M+Na]^+$, $[M+K]^+$, $[M+NH_4]^+$ or $[M-Cl]^-$
- Both positive and negative ion data are typically collected

Electrospray Ionisation (ESI) - Theory

As the eluent passes through the capillary the electric field disrupts the emerging liquid surface

Eluent protrudes from the end of the capillary as a Taylor cone

+ve potential on capillary → +ve ions
-ve potential on capillary → -ve ions



High potential (1-4 kV) is applied between the tip of the capillary and a counter electrode

The droplets move towards the orifice of the reduced pressure regions of the mass spectrometer

Monoisotopic mass

Molecules are typically measured in the ground-state of the principal (most abundant) isotope for each element (monoisotopic mass)

Example cortisol:

- Formula: $C_{21}H_{30}O_5$
- Molecular mass: 362.460 Da g/mol (average)
- The corresponding monoisotopic mass is:
362.209320 Da g/mol

Naturally occurring isotopes

Isotope	Natural Abundance (%)	Isotopic mass (u)	Isotope	Natural Abundance (%)	Isotopic mass (u)
¹ H	99.985	1.007 825	¹⁹ F	100	18.998 405
² H	0.015	2.014 102	²³ Na	100	22.989 767
¹² C	98.9	12.000 000	³¹ P	100	30.973 763
¹³ C	1.1	13.00 3354	³⁹ K	93.3	38.9637074
¹⁴ N	99.64	14.00 3074	⁴¹ K	6.7	40.961825
¹⁵ N	0.36	15.000 108	³⁵ Cl	75.8	34.968 855
¹⁶ O	99.8	15.994 915	³⁷ Cl	24.2	36.965 896
¹⁷ O	0.04	16.999 133	⁷⁹ Br	50.5	78.918 348
¹⁸ O	0.2	17.999 160	⁸¹ Br	49.5	80.916 344

The mass of ¹²C is exactly 12.000000 u,
 u (unified atomic mass) = 1/12 of the mass of an atom of ¹²C

Mass analyser

Many different instrument types and configurations:

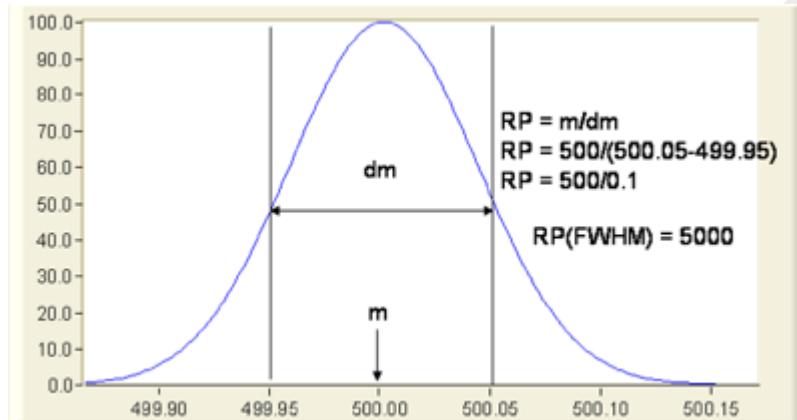
- Low mass resolution, e.g. quadrupole
- High mass resolution, e.g. quadrupole-time-of-flight (QToF), Orbitrap
- Untargeted or targeted approach?
 - Untargeted approaches are exploratory and rely on high resolution accurate mass instruments for the identification of features of interest
 - A targeted approach requires you to know what you want to measure
 - Principally based on nominal mass instruments using MS/MS methods
 - up till 2015, typically used in routine laboratories for assays
 - Now we also use high mass resolution instruments in the clinic

Mass resolution

Mass resolving power = ability to separate two mass spectral peaks

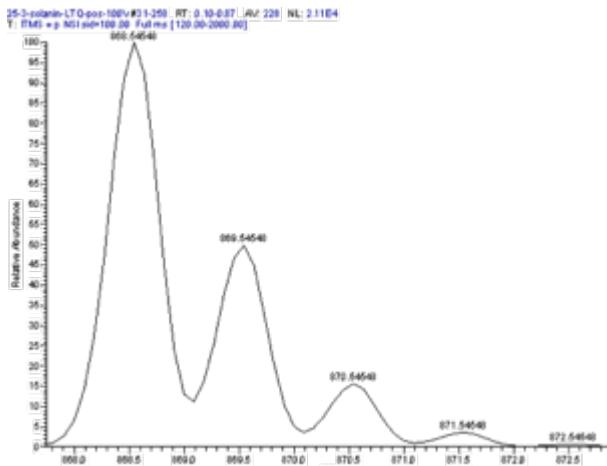
- Defined as m/dm

where m designates the mass and dm the peak width necessary for separation of mass m . A specific m/z value and also the method (typically 50% valley of full width at half maximum (FWHM) must be given).



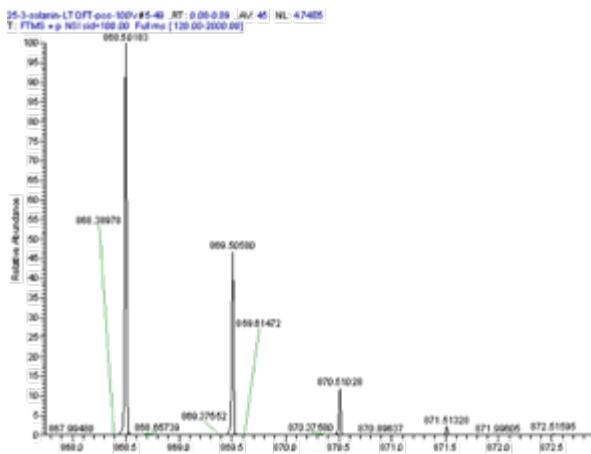
Mass resolution

Low resolution



LTQ, resolving power R = 1737 at m/z 868.5, peak width ~ 0.5 FWHM

High resolution



LTQ-FT, resolving power R = 48,250 at m/z 868.5, peak width ~ 0.018 FWHM

Mass accuracy

- Mass accuracy = degree of conformity of a measured quantity to its actual value
 - Typically reported in ppm:

$$ppm = \left(\frac{m_{exp} - m_{obs}}{m_{obs}} \right) \times 10^6$$

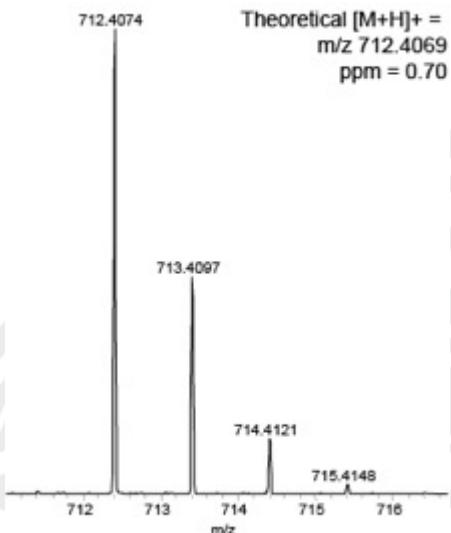
where m_{exp} = theoretical mass and m_{obs} is the measured mass

e.g. Theoretic mass = 500.0025
 Measured mass = 500.0000

$$ppm = \left(\frac{500.0025 - 500.0000}{500.0000} \right) \times 10^6 = 5 \text{ ppm}$$

Mass accuracy

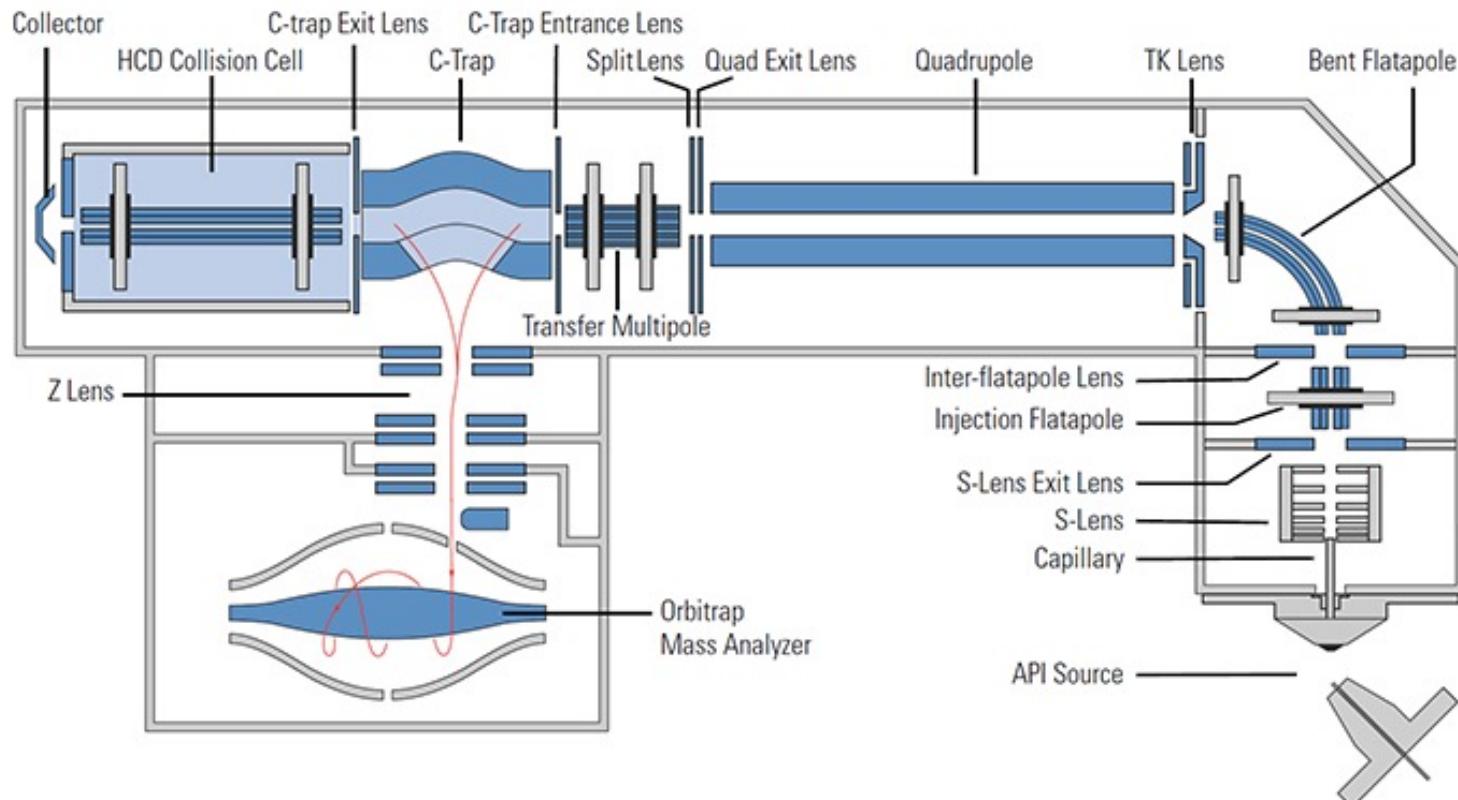
- Accurate mass measurements take advantage of the fact the the combination of elements contained in a molecule actually have a very specific, non-nominal molecular weight
 - Carbon has a mass of 12.0000
 - Hydrogen has mass of 1.0078
 - Oxygen has a mass of 15.9949
 - Nitrogen has a mass of 14.0031
- It is possible to have combination of atoms which have the same nominal (integer) mass but different accuracy mass
- If such compounds can be measured with sufficient mass accuracy it is possible to determine the elemental composition



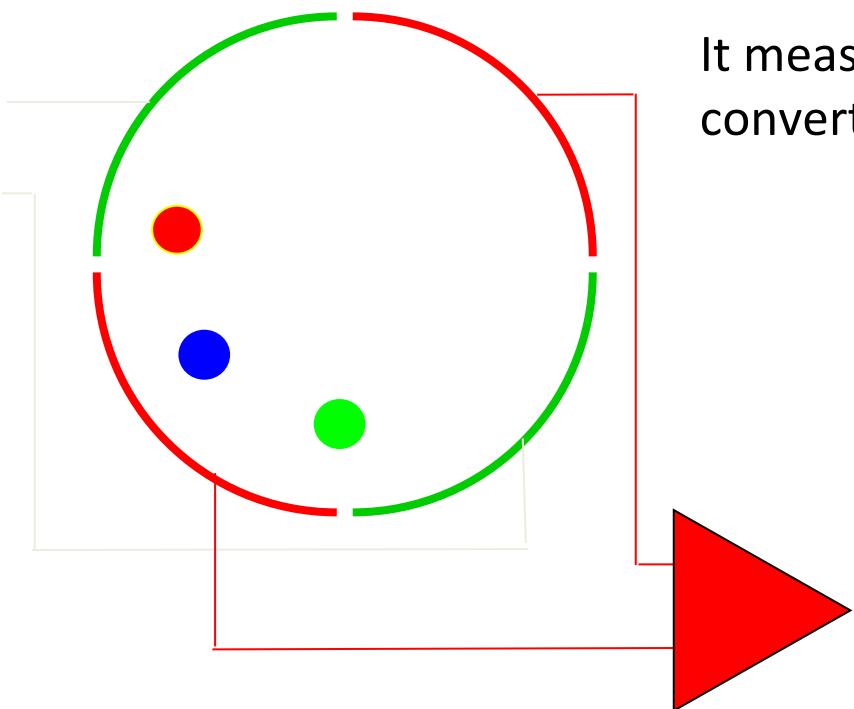
High resolution mass analysers

- For untargeted analysis it is important to have high mass **resolution, accuracy** and **speed**
- When high resolving power is needed:
 - Time-of-Flight (ToF)
 - High resolution ion traps
 - Orbitrap
 - Fourier transform Ion Cyclotron Resonance (FT/ICR)

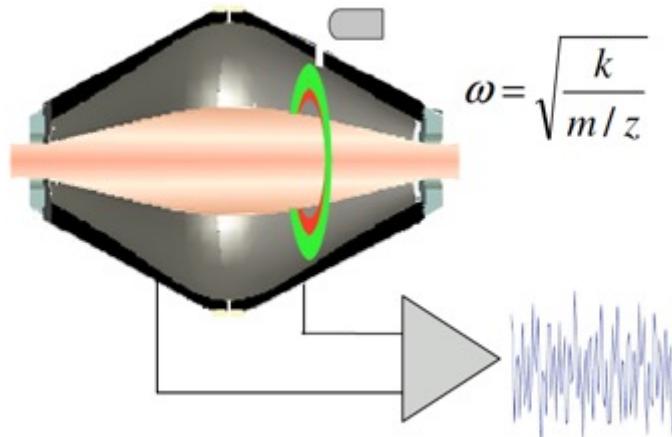
High resolution Orbitrap



The Orbitrap is not a scale



It measures oscillations that can be mathematically converted to mass/charge (m/z)



Fourier transformation of data to produce m/z

Resolution & Mass Accuracy

Type	Resolving Power (FWHM)	Mass Accuracy (ppm)
FT-ICR-MS	1,000,000	0.1 - 1
Orbitrap	100,000	0.5 - 1
High-Res-TOF	60,000	3 - 5
ToF	10,000	3 - 5
Triple Quadrupole	1,000	3 - 5
Ion Trap	1,000	50 - 200

Uniqueness of molecular ions

- The molecular ion, even when measured with the highest accuracy, is not a unique descriptor
- There are many theoretical possible structures for a given mass and empirical formula

Searched arginine m/z: 174.1116 with 10 ppm tolerance

 HMDB [Browse](#) ▾ [Search](#) ▾ [Downloads](#) [About](#) ▾ [Contact Us](#)

			Search	metabolites	Search	
HMDB0041906	Ibopamine	C17H25NO4	307.1784	M+ACN+2H	175.1097	2
HMDB0034696	Artemin	C15H22O4	266.1518	M+2ACN+2H	175.1097	2
HMDB0030104	Humulinic acid A	C15H22O4	266.1518	M+2ACN+2H	175.1097	2
HMDB0036151	Arabsin	C15H22O4	266.1518	M+2ACN+2H	175.1097	2
HMDB0037059	1alpha-Hydroxyarbusculin A	C15H22O4	266.1518	M+2ACN+2H	175.1097	2

Showing 1 to 10 of 166 entries

Previous 1 2 3 4 5 ... 17 Next

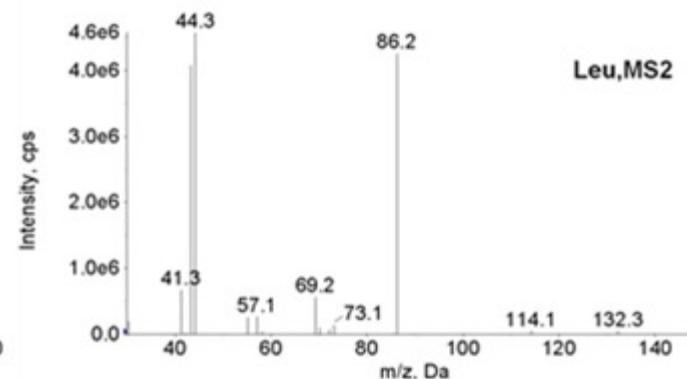
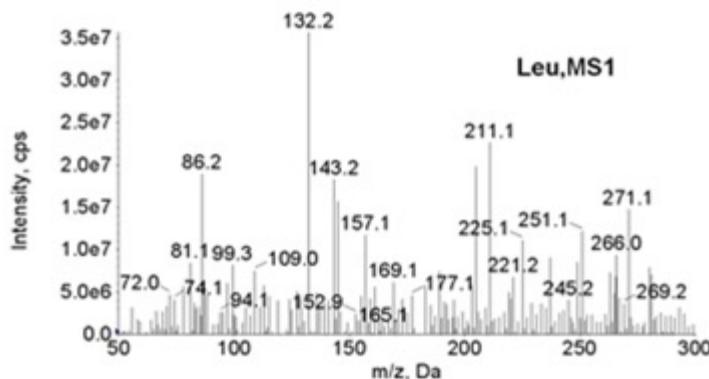
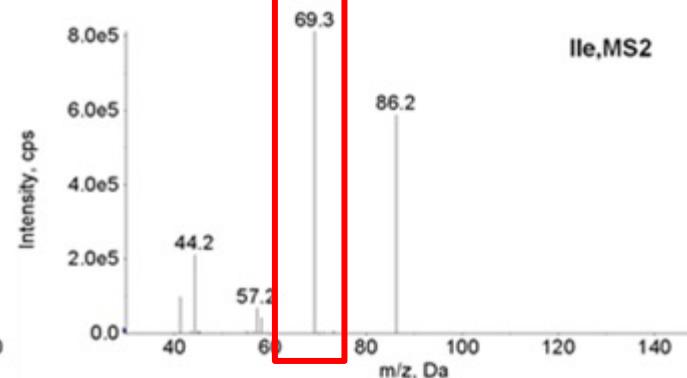
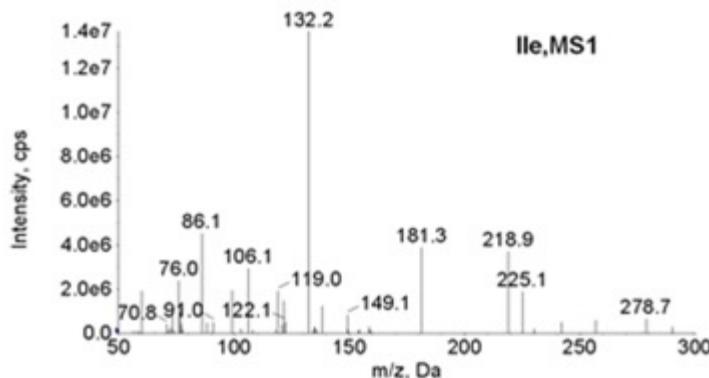
Tandem MS/MS

- A second mass spectrum (MS/MS) that is informative arises from isolating the molecular ion
- The molecular ion is heated, either by collision with neutral gas (collision induced dissociation (CID) – quadrupole, ion traps) or Higher-energy collisional dissociation (HCD) (Orbitraps)
- The extra energy increases the stretching of critical bonds, leading to dissociation of the molecular precursors ion into charged product ions
 - These generate the MS/MS spectrum for a metabolite
 - Ion traps can also isolate a product ion and create MS_n spectra

Fragmentation (MS/MS)

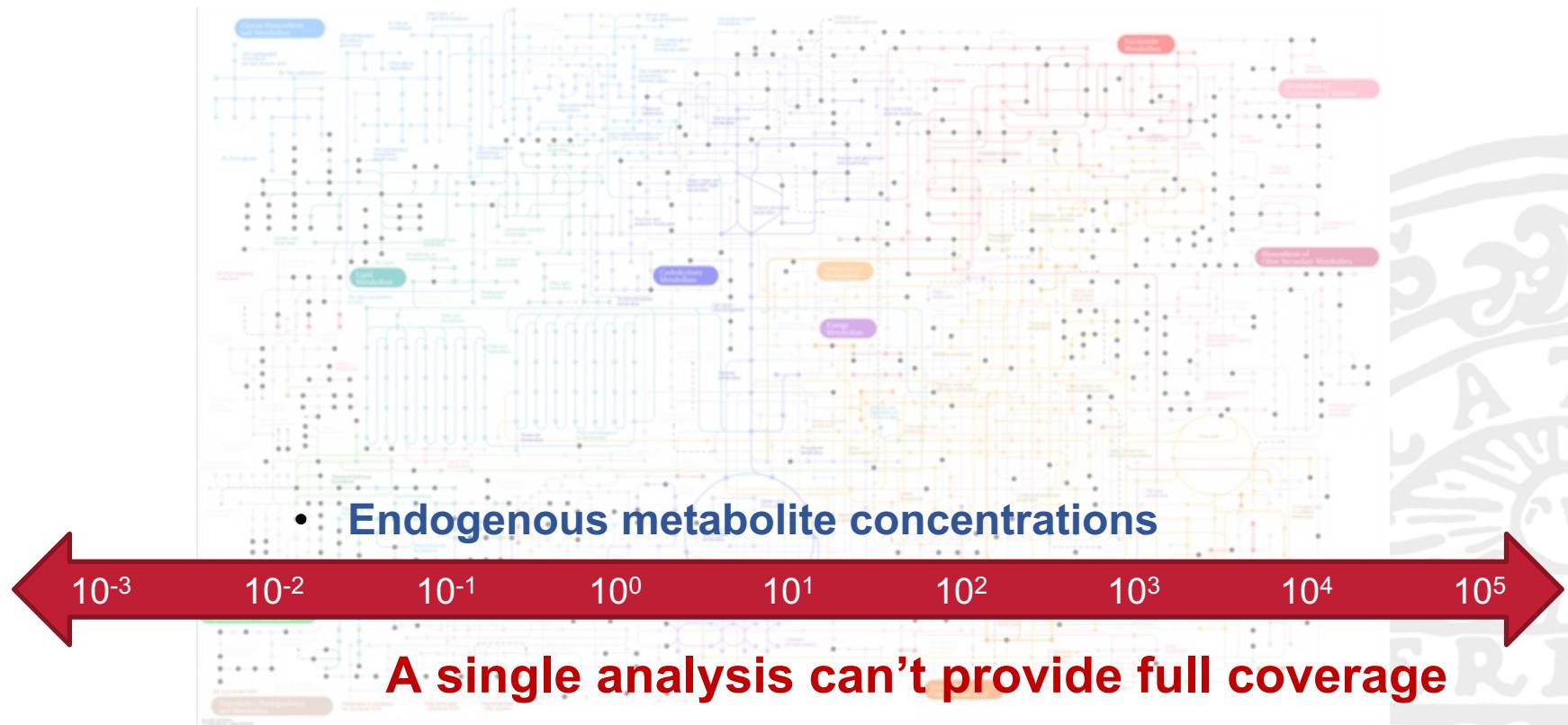
leucine and isoleucine are isobaric compounds
- Exact same mass

Unique fragment pattern
enables discrimination between
leucine and isoleucine



Challenges in Metabolomics

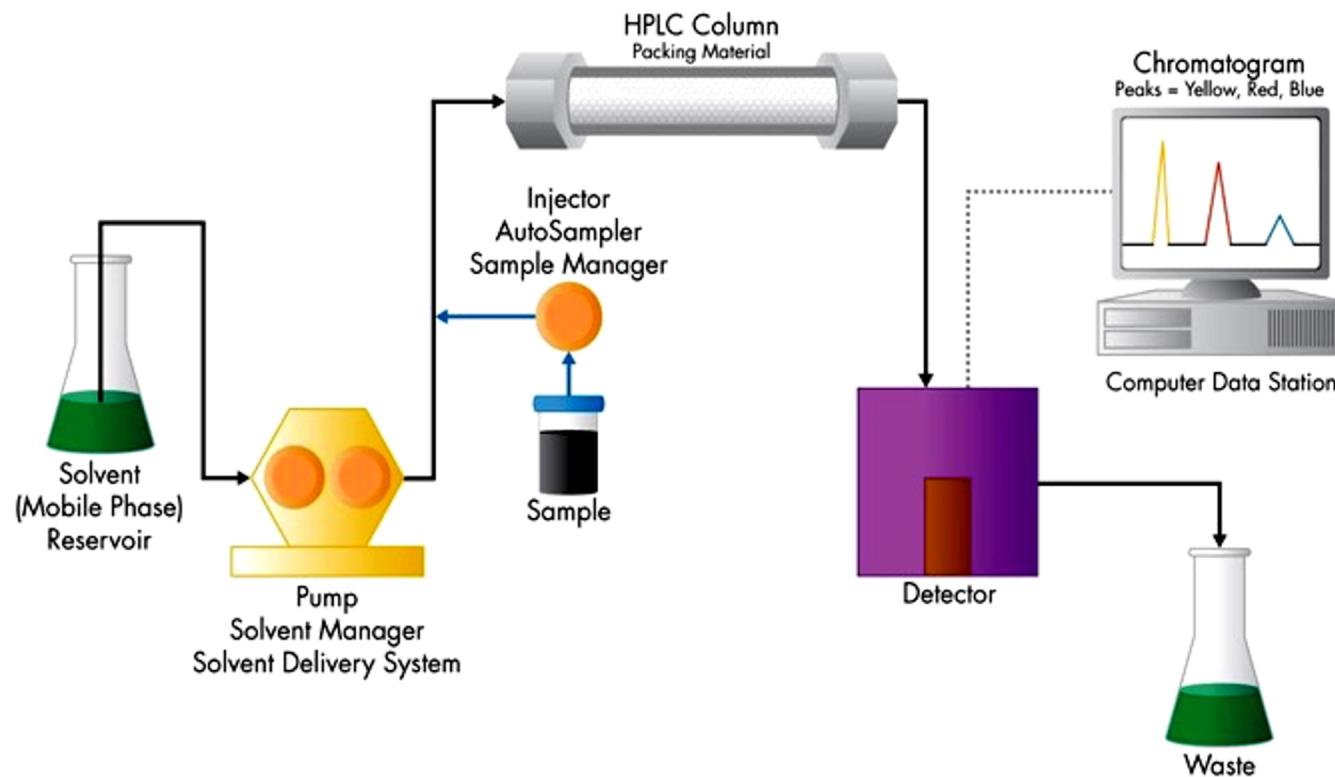
Differences in physiochemical properties among metabolites



Direct infusion MS (DIMS)

- High-throughput screening
- Crude sample extracts injected or infused into the mass spectrometer
- Coverage depends on ability of the metabolite to be ionized
- Ion suppression
- Not quantitative
- Unable to distinguish isomers

Separation of compounds using Liquid Chromatography (LC)



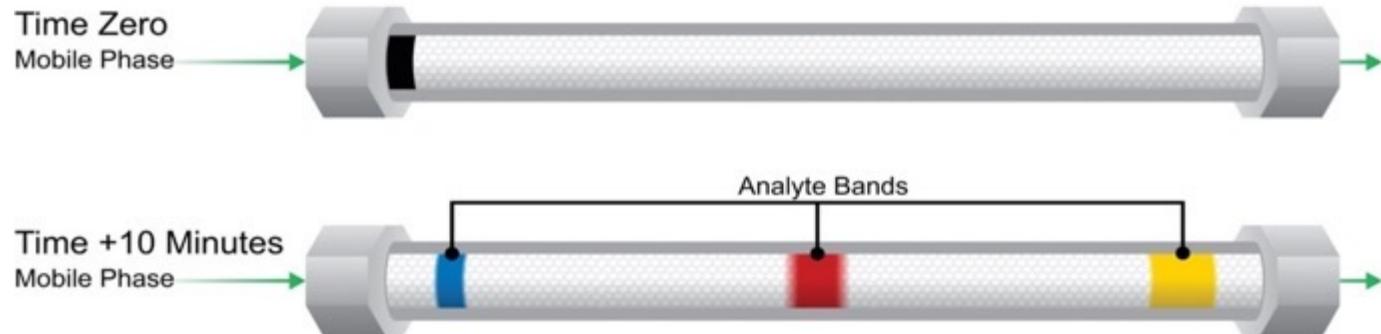
Why perform separation?

- To add a second dimension to MS analysis
 - Maximising specificity (separating molecular species of similar or identical mass)
- To minimise ionisation suppression and maximise intensity
 - Higher quantitative accuracy
- Complementary techniques
 - Liquid chromatography (LC)
 - Gas chromatography (GC)
 - Capillary electrophoresis (CE)

Liquid Chromatographic Separation

We create a separation by changing the ***relative speed*** of each analyte band
(competition between the mobile phase and stationary phase)

Injected Sample Band (Appears “Black”) (Blue, Red, Yellow)



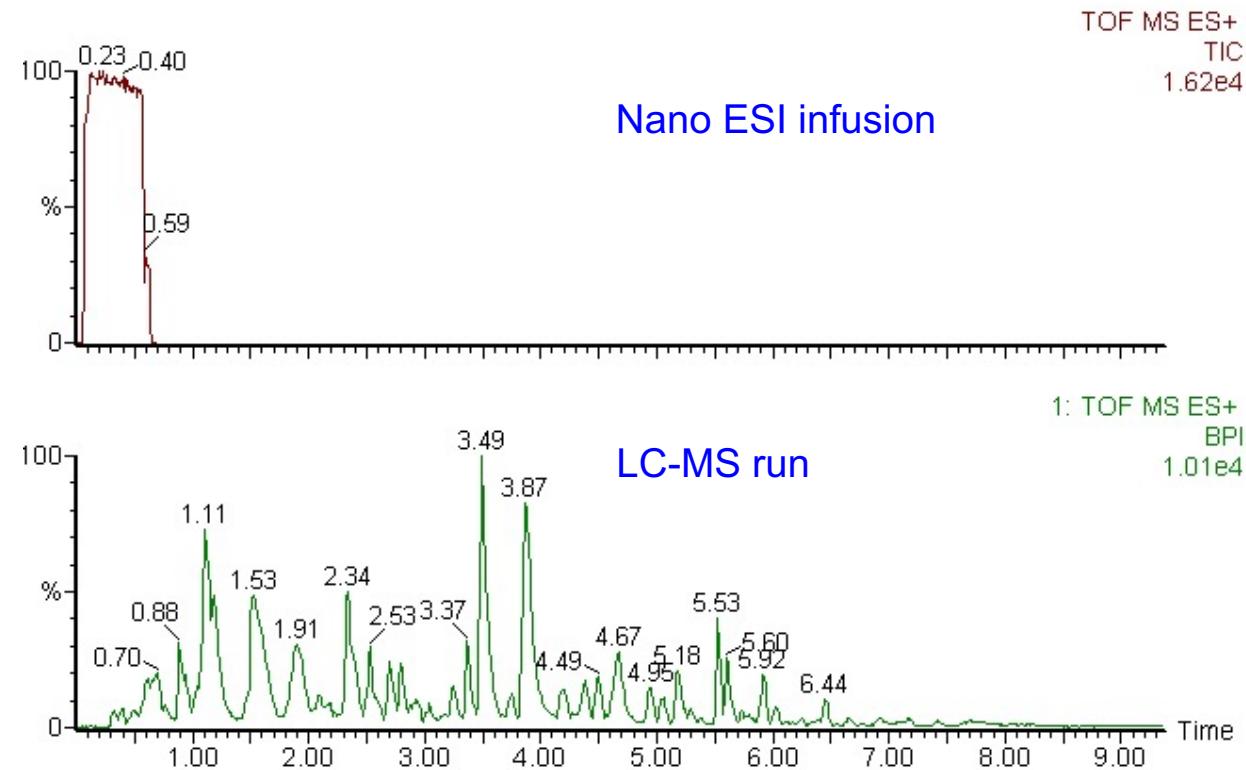
Yellow is the earliest eluting analyte “band” (it will be broader in the column), but moving fastest – it “likes” the mobile phase

Blue is well retained, it will be in a more focused, narrower band, near the inlet and move the slowest in the column – it “likes” the particles

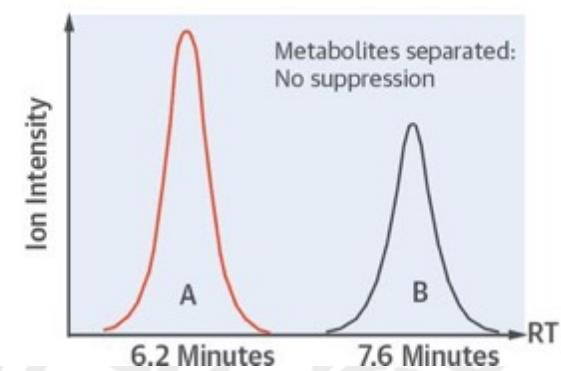
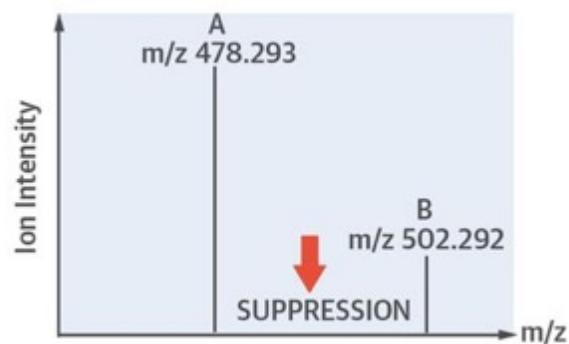
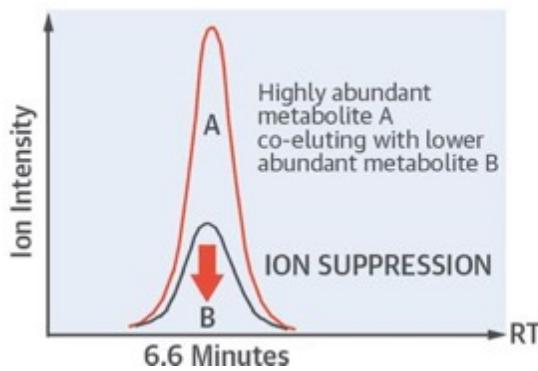
Courtesy of Waters® Corporation

Direct infusion vs. LC-separation

A comparison between NanoESI infusion vs LC for a rat urine sample



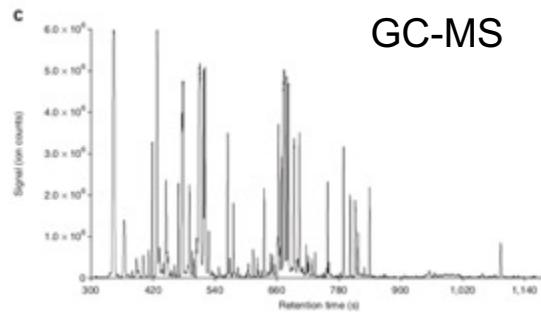
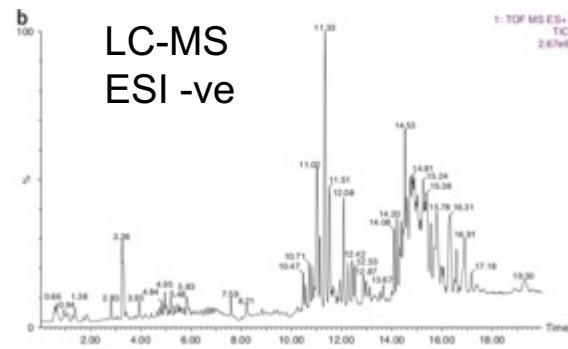
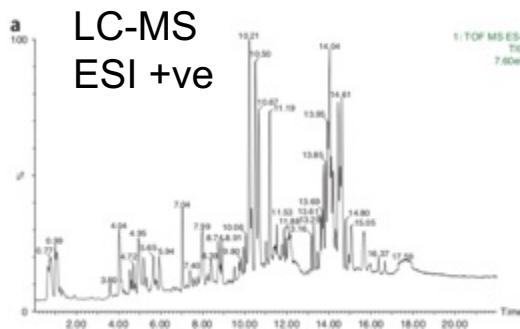
Separation- less Ion Suppression



Baig F et al. JACC, 68 (2016) 1294

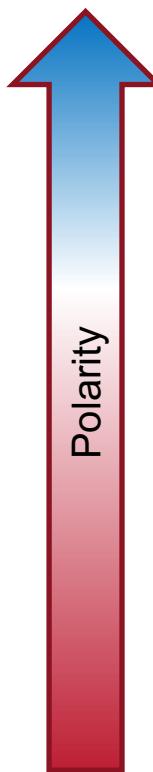
What you choose will determine what you see

Typical chromatograms observed for serum



Complex sample – separation of metabolites needs different separation chemistry

Purines & pyrimidines
Sugars
Organic acids
Amino acids, amines
Bilirubin
Bile acids
Eicosanoids & metabolites
Fatty acids
Other phospholipids
Phosphatidylethanolamines
Phosphatidylcholines
Sphingomyelins
Diacylglycerols
Cholesterol esters
Triglycerides



HILIC as a separation technique is the strong retention of polar, hydrophilic compounds



Reversed-phase C18



Reversed-phase C8

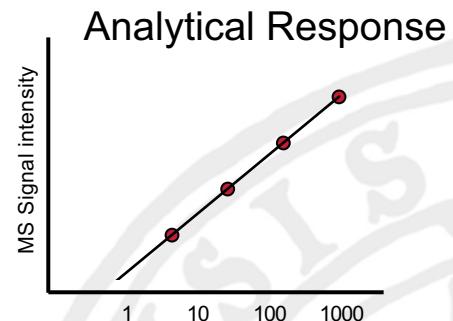
Reversed-phase as a separation technique is the strong retention of nonpolar, hydrophobic compounds

Mass spectrometry based metabolomics

Sensitivity dependent on analyte and ionisation



Untargeted profiling
100s -1000s
metabolites
Typical measurement
over 4 orders of
magnitude



Endogenous metabolite concentrations



A single analysis can't provide full coverage

Big Data in metabolomics- Challenges

- The rapid progress in field has resulted in a mosaic of independent, and sometimes incompatible, analysis methods that are difficult to connect into a useful and complete data analysis solution
- Configuring necessary software tools and chaining them together into a complete re-runnable analysis

Big Data in metabolomics- Challenges

File formats, all vendors have their own...

The mzML format is an open, XML-based format for mass spectrometer output files, developed with the full participation of vendors and researchers in order to create a single open format that would be supported by all softwares

Concluding remarks- What to choose?

- Type of sample
- Type of analytes
- Degree of separation required
- Analysis time
 - If you want the maximum number of markers then use longer runs
- What platforms are available?