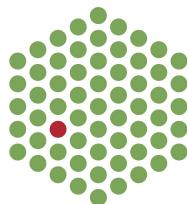


Untargeted Metabolomics workflow: PSI Metabolomics Workshop 2019

Prasad Phapale
EMBL Heidelberg, Germany

NDRI Karnal, India

EMBL



Metabolomics

Relevance to biological response/ Phenotype

WHAT RESPONSES ARE POSSIBLE

GENOMICS



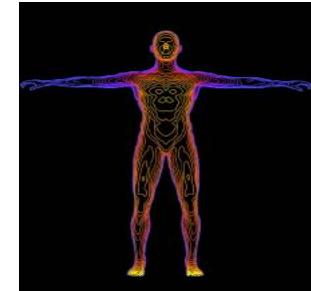
INITIATION OF RESPONSE

TRANSCRIPTOMICS



Environment

Functional Metabolomics



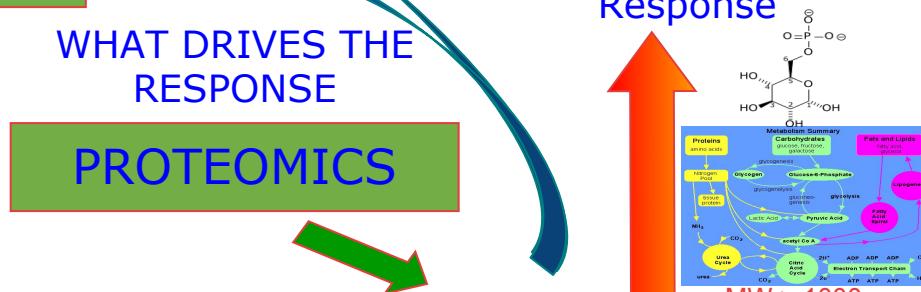
PHENOTYPE/
Drug or disease
Response

WHAT DRIVES THE RESPONSE

PROTEOMICS

THE ACTUAL RESPONSE

METABOLOMICS

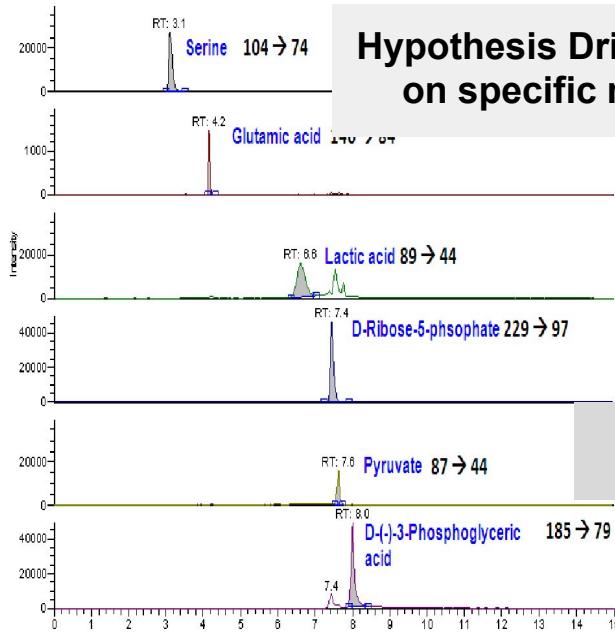


Metabolomics vs Proteomics

- Metabolite analysis fundamentally different than protein/ peptides
- Sample prep: protein removal
- Chromatography
 - Shorter gradient
 - High flow
 - Different mobile phase buffers and columns
- Lower mz range; negative mode
- Identification and annotations:
 - Diverse chemistry

Metabolomic approaches

Targeted Metabolomics

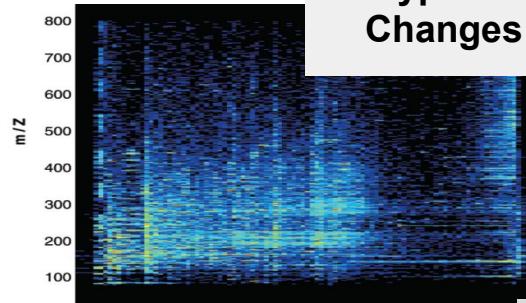


Biomarker quantification

Hypothesis Driven: Focused on specific metabolites

Triple quad LC-MS/MS

Untargeted Metabolomics

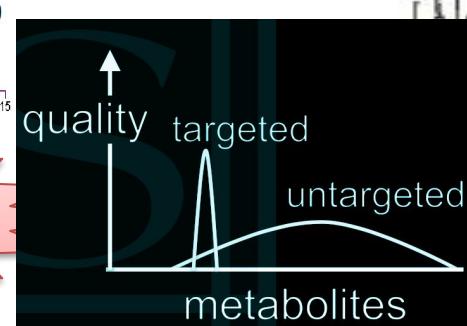


Hypothesis Generating: Changes in metabolome

LC-HRMS/MS



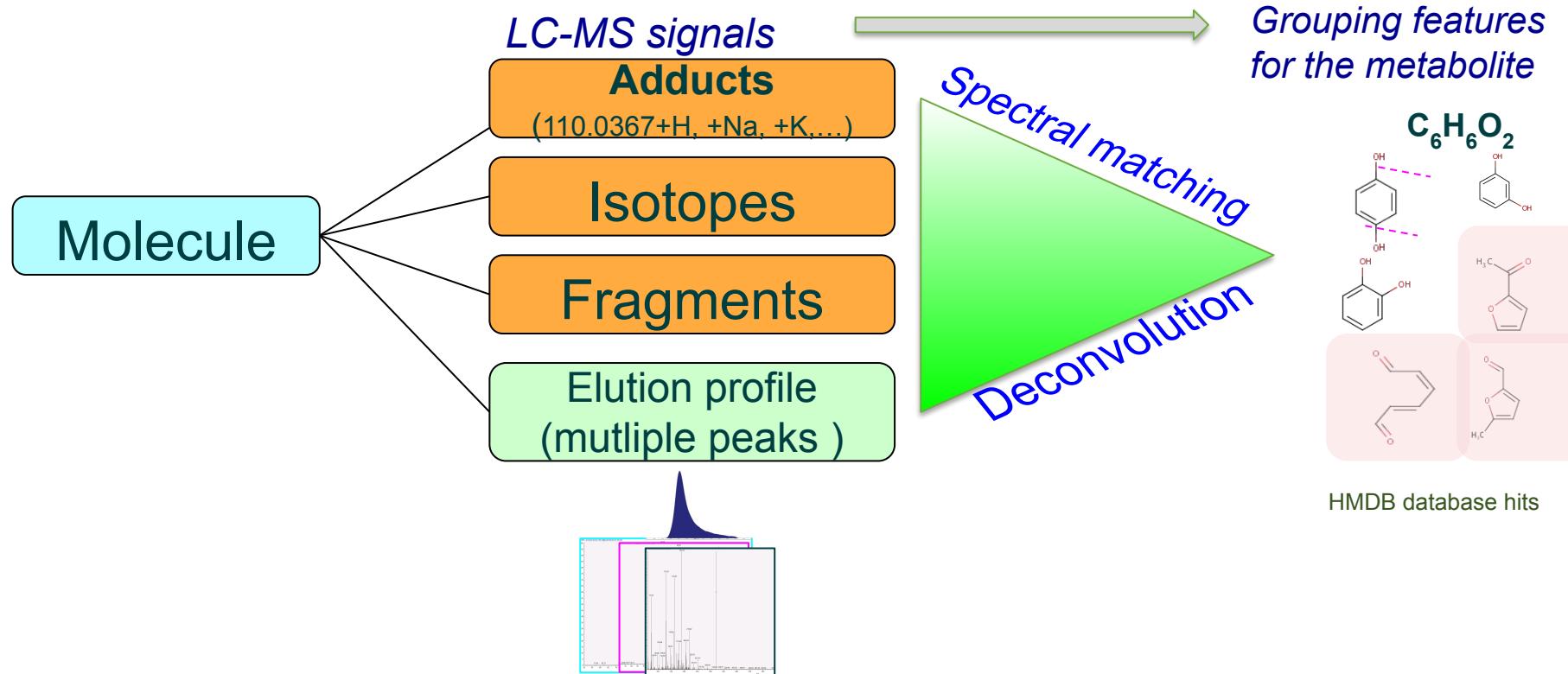
Biomarker discovery



Untargeted not Global

- A Metabolomics Method really measure thousands of unique metabolites in a single analysis?
 - No
 - peaks/ feature  metabolites
- What are peaks?
 - Adducts e.g. $[M+H]^+$, $[M+Na]^+$, $[M+K]^+$, $[M+NH_4]^+$, etc.
 - Fragments
 - Isotopes
 - Dimers, trimers
 - Multiple chromatographic peaks; separation issues
 - Contaminants
 - Software algorithm errors: missing peaks, peak integration, noise
 - Chemical noise
 - Instrumental noise

Annotation Challenge



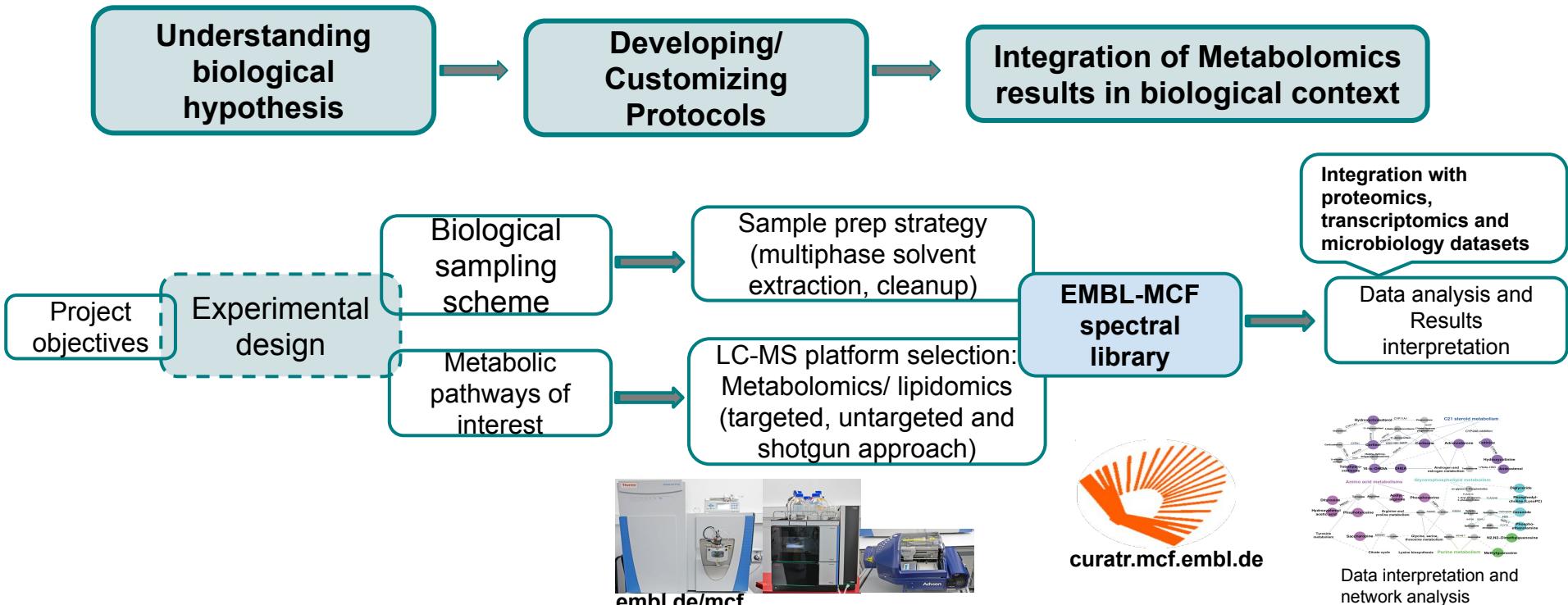
Why LC ESI-MS for Metabolomics?



Untargeted is challenging

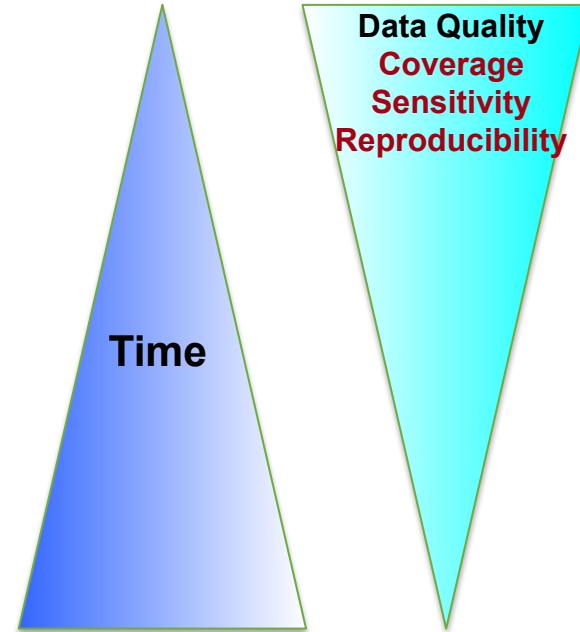
- Have hypothesis to test?
- Are alternative experiments or targeted approach possible?
 - Probably better
- If other designs are not possible
 - Plan Untargeted experiments carefully
- **Semi-targeted approach**
 - Targeted pathways
 - Optimize suitable LC-MS platform
 - In-house standard library

Workflow management at EMBL



Workflow

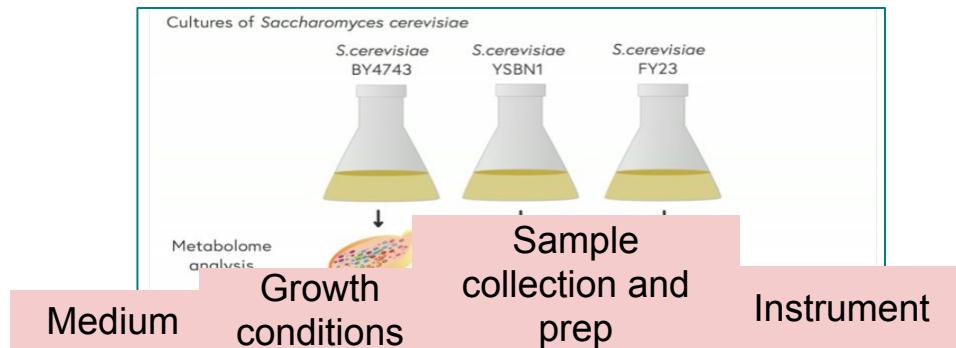
- Experimental design
 - controls (pos/ neg)
 - replicates
- Sample collection
 - stability
 - time
- Metabolite Extraction
 - Class specific
 - Clean-up
- Separation and Detection
 - LC-MS method
- Data analysis
 - Annotation/ IDs (non-targeted)
 - Quantification



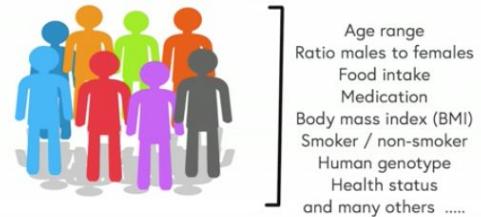
Experimental Design

Aim of Exp Design

- Controlling all other sources of variation while capturing biological variation from samples
- Controls: to track variations
 - Biological
 - Analytical
- Replicates
 - Controlled lab studies < Uncontrolled clinical studies more



Clinical studies: different aspects

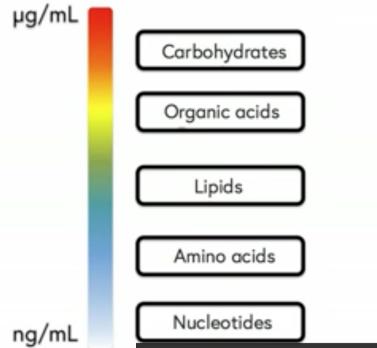


Control groups

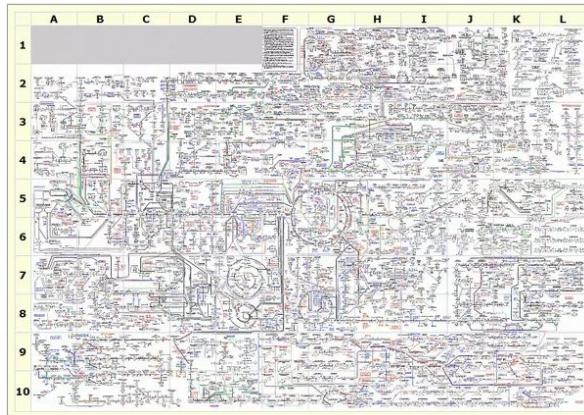
Considerations for semi-targeted approach

Coverage, sensitivity → chemical diversity

Dynamic Range of Metabolites



Not possible to detect all metabolites on one platform
(Untargeted metabolomics)



- Selection of
 - sample collection and prep method based on sample types
 - Analysis platforms based on targeted pathways/ class

Quantitative Untargeted Metabolomics

Sample amount

Metabolite extraction

Instrumental analysis (LC-MS)

Considerations

Experimental design, sample type, collection method, stability

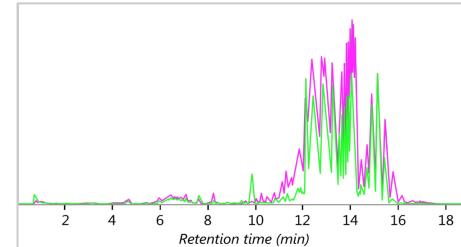
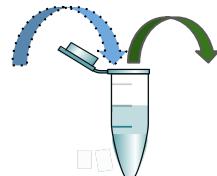
Cell numbers, weight, volume, biomass, protein/DNA concentration

Extraction efficiency, recovery, quenching

Internal standards

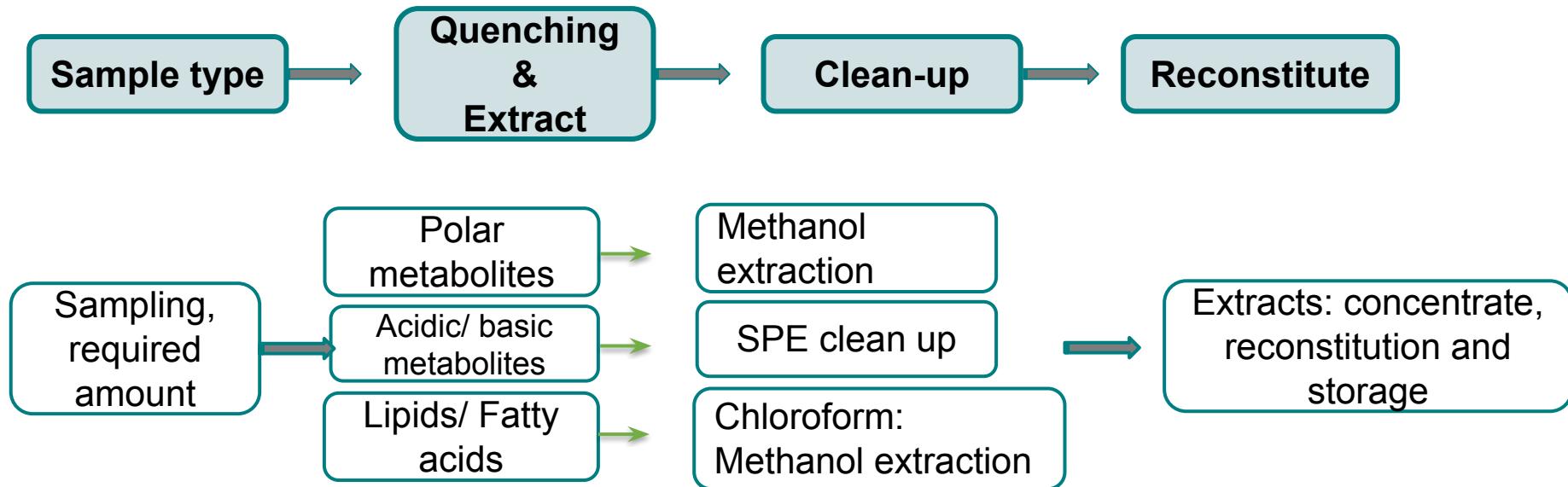
Reproducibility, sensitivity

Quality controls



Biological Samples

Workflow: Sample prep



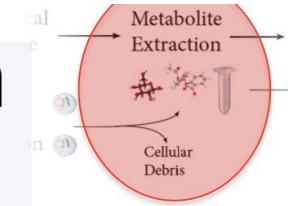
Metabolite Extraction

- Goal

- Separation of metabolites from cellular debris
- Protein, DNA, RNA precipitation
- Dissolve both polar and non-polar metabolites in solution
- Minimal losses due to degradation or biochemical conversion

- Practically

- Bias and losses are inevitable due to a wide range in solubility of this diverse group of compounds
- Extraction often results in a dilute extract

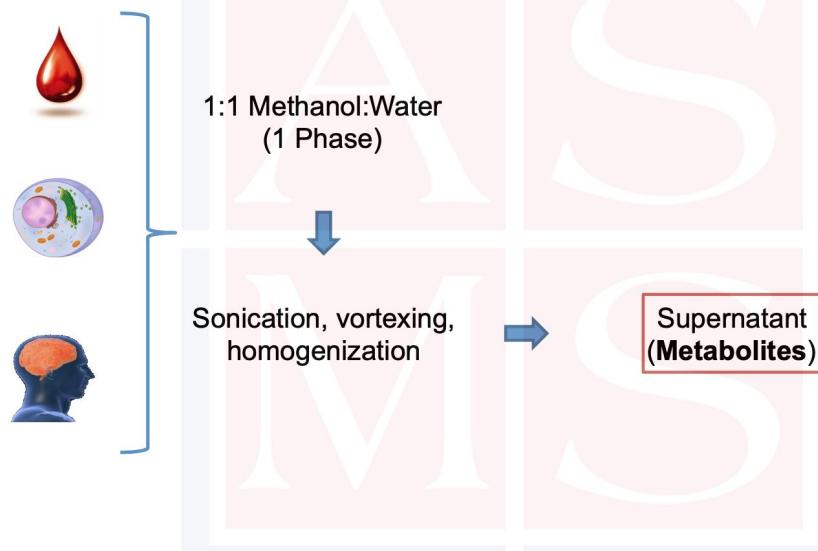


Parameters to consider:

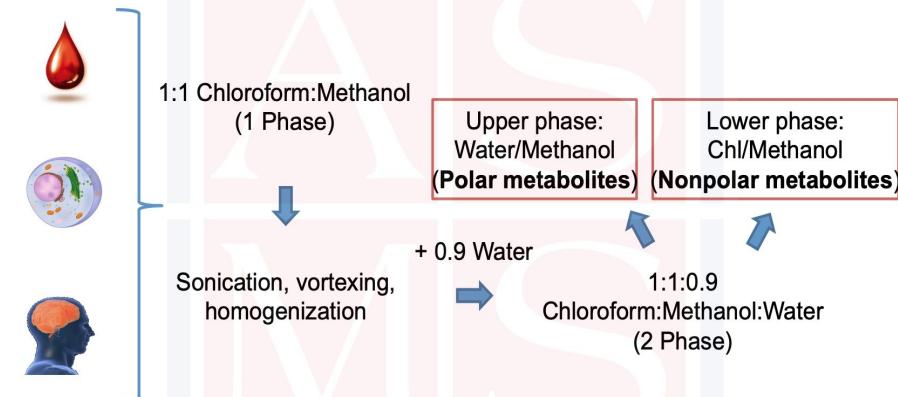
1. Solvent
2. pH
3. Temperature
4. Multiple extractions
5. Mol wt cut-off filters

*Adapted From ASMS 2018
Workshop by Gary Patti*

Monophasic Extractions



Biphasic Extractions



- Increasingly common.
- Multiple analyses on different column chemistries.

Adapted From ASMS 2018
Workshop by Gary Patti

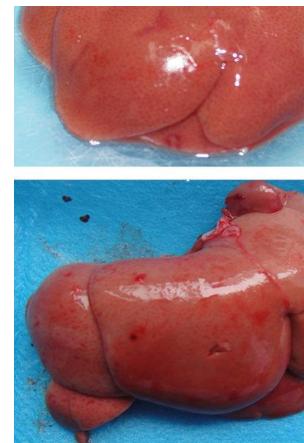
Sample specific considerations

- Cell washing
- Quenching
- Suspension vs Adherent



Cells

- Collection, storage
- Weight
- Homogenization



Tissue

- Collection, storage
- Aliquots
- Processing, throughput



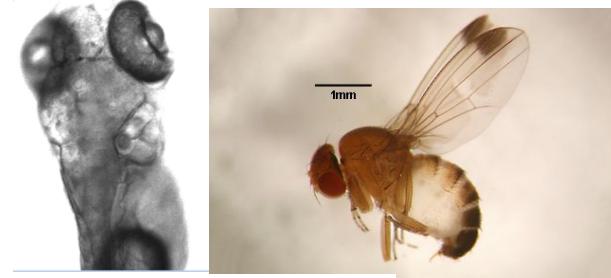
Biofluid

Sample types



Bacterial agar plates
Microbial co-cultures

Microbial cultures



Model organisms



Sampling,
treatment specific



Homogenization

Plant tissues

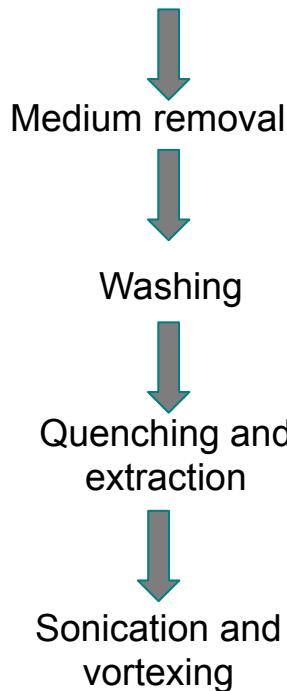
Orthogonal Measurements

- Collect orthogonal data
 - Protein conc
 - Cell biology
 - Enzyme assays
- Clinical
 - Meta-data (age, medications, diet, etc.)
 - Sampling conditions
 - Pathological analysis

Cell culture

Mammalian Cell culture

Adherent cells



Suspension cells



Cells and medium

- Count cell numbers
- Measure protein concentrations from cell pellet
- Measure DNA concentrations
→ use for data normalization!

Medium removal

- Extracellular metabolome
- Medium content
- Serum precipitation with 6x solvent
- Concentrate
- Fresh medium control



Washing

- Gentle washing with suitable buffer
- PBS may not be the best
- Preferably use volatile buffer (formate, acetate)
 - Check if your cells are OK!
 - Watch for cell response/ damage
- Optimize buffer concentration !

Quenching

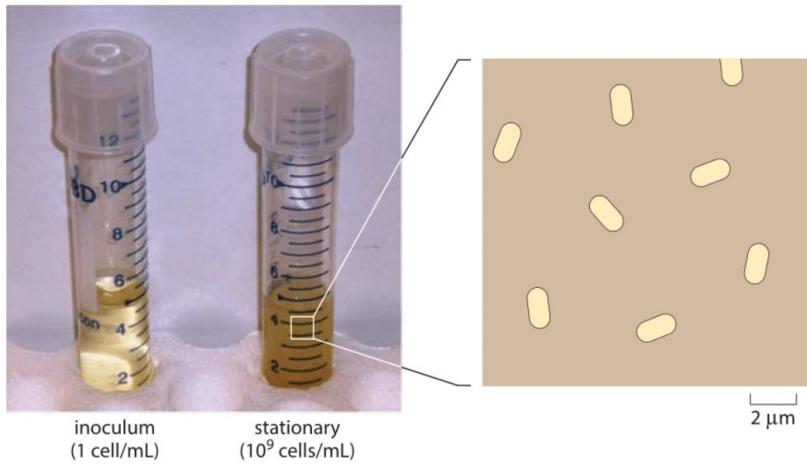
- Cold solvent (Methanol)
 - Consistent and quick
- Liquid nitrogen
 - Fast, volatile,
no-homogeneous
- Hot solvent
 - Efficient
 - degradation
- Acid/ alkali
 - Fast and effective
 - Less suitable for some
metabolites

	E coli	HeLa cells
Cellular pool half-life (dominated by)		
1 s	Metabolite (turnover)	1 min
10 min	mRNA (degradation)	10 hr
1 hr	Protein (dilution)	1 day

SnapShot: Timescales in Cell Biology. Cell (2016)

Avoid trypsinization of cells!

Bacterial culture



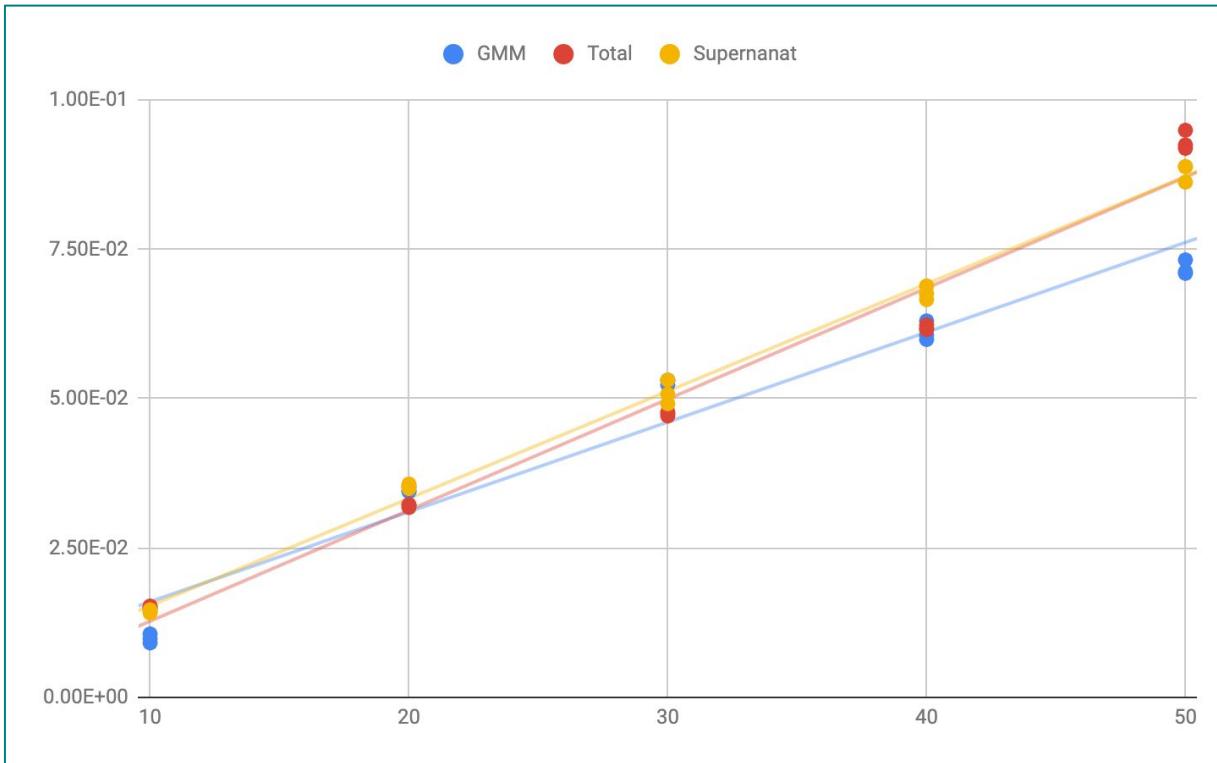
organism	(cells/mL)/OD ₆₀₀	CDW/OD (g/L)/OD ₆₀₀	yield (g CDW/g sugar)	specific cell volume (mL/g CDW)
<i>E. coli</i>	$0.6-2 \times 10^9$	0.3-0.5	0.17-0.55	1.3-2.8
<i>S. cerevisiae</i>	$0.8-3 \times 10^7$	0.5-0.6	0.45	2

Table 1: Conversion between optical density (OD) and cell concentrations. CDW is cell dry weight.

- Separation of medium and cells
- Matrix effect GMM medium

Source: Cell biology by numbers
<http://book.bionumbers.org/>

Bacterial culture: metabolite analysis



**Consider matrix effect
when comparing data
from spent medium, total
and cell pellet**

Yeast culture

- Filtration
- Mechanical cell lysis in cold solvent (eg; glass beads, homogenizer)
- Freeze-thaw cycles

Avoid: detergents, phosphates, reagents
→ need further clean-up

Tissue

- Frozen tissues
- Dry weight (5 to 25 mg)
- Homogenization: mechanical
- Extraction
- Minimize freeze-thaw cycles, warm time (storage)
- Maximise surface area

Drying/ concentration of extracts

- Speed-vac
 - Warming effect!
- Nitrogen gas evaporator
- Complete drying is not required
 - Should be uniform

Reconstitution

- Compatible with mobile phase
- Compatible with extracts
- ACN can cause phase separation in some cases
 - Check visually
 - Use methanol
- Critical for chromatography, peak shape and consistency
- Centrifuge before injection
 - Check for consistency

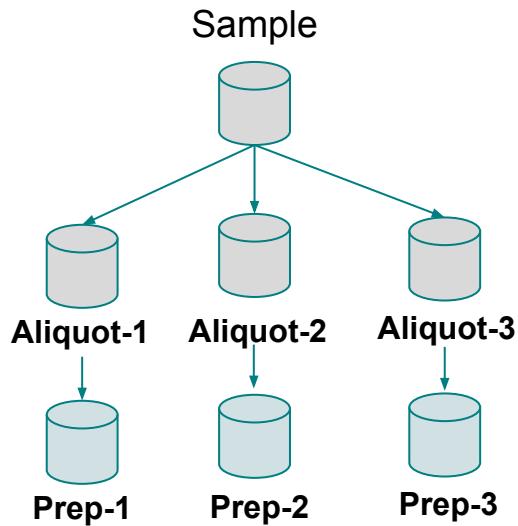
Replicates

Replicates: Biological

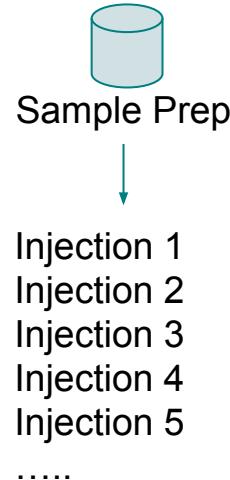
- Depends on variability between samples
 - Animal studies < Human samples
- Purpose of the experiment
- power analysis in Metaboanalyst (*Chong et al., 2018*)
- 3 to 5 replicates is good number to start
- For large studies 30 per group
- Better stat and confident results

Replicates: Technical

Preparations Replicates (aliquots)



Injection Replicates (instrument)



- For initial validation
- For routine analysis use QC approach

Chromatography

- Slow
 - Faster than proteomics (run times 8 to 30 min)
- Additional dimension to the data
 - RT
- Reduces matrix effect
 - Better quantification
- Quality MS/MS data

Stationary phases

- RP
 - Robust
 - May not be the best option for imp metabolites
- HILIC
 - Longer stabilisation time; less stable, RT variation
 - retention for polar metabolites (AA, glycolytic,etc)
- RP-ion pairing
 - Good retention
 - MS contamination! → need dedicated MS!
- Mixed mode
 - Ion exchange, Acclaim trinity P1, diamond hydride, Mastro C18

HILIC

- For glycolytic and central carbon metabolites
- More coverage high pH upto 9
 - NH₂ column best; shorter life
 - Amide column better option
- High organic solvent: efficient ESI ionization/ response
- Long equilibration time
- Hydrophobic metabolites-> less retention
- Column bleeding issue at high pH

Which mode?

- HILIC
- RP
- Mix-mode

Multiplex or use chromatography best for your samples

Without chromatography

- MALDI
 - Matrix application critical
 - Laser ionization; more suitable for lipids
 - Quantitation challenges
- Nanomate (shotgun)
 - Sensitive, fast
 - Spray stability issue
 - Can be coupled with LC
- Flow-injection
 - High-throughput
 - Matrix effect, artifacts

Contaminations!

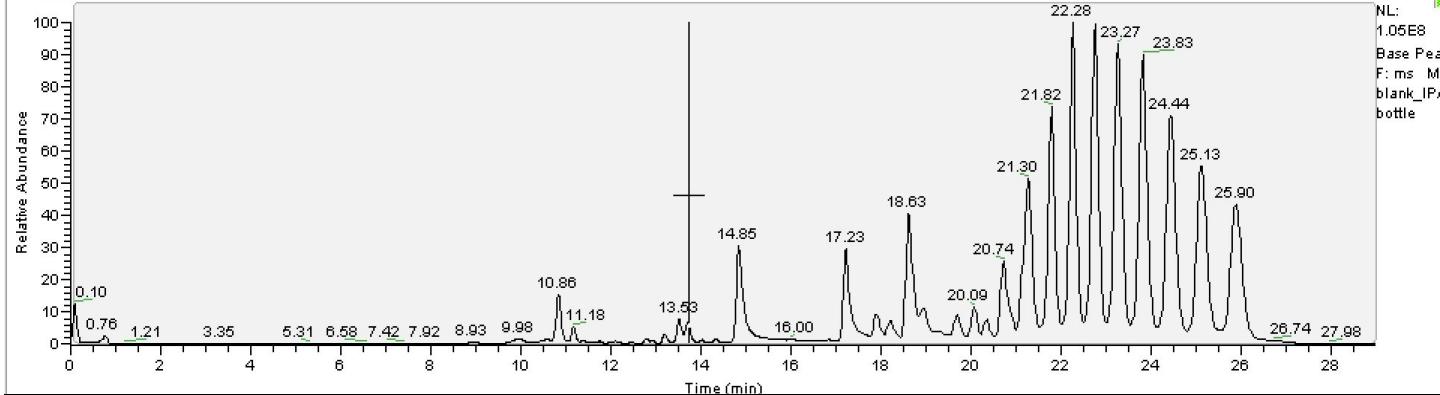
Blanks

- Solvent or mobile phase blank
- Instrument blank (0 uL)
- Processed blank
 - Processed same as samples
 - Background from sample processing

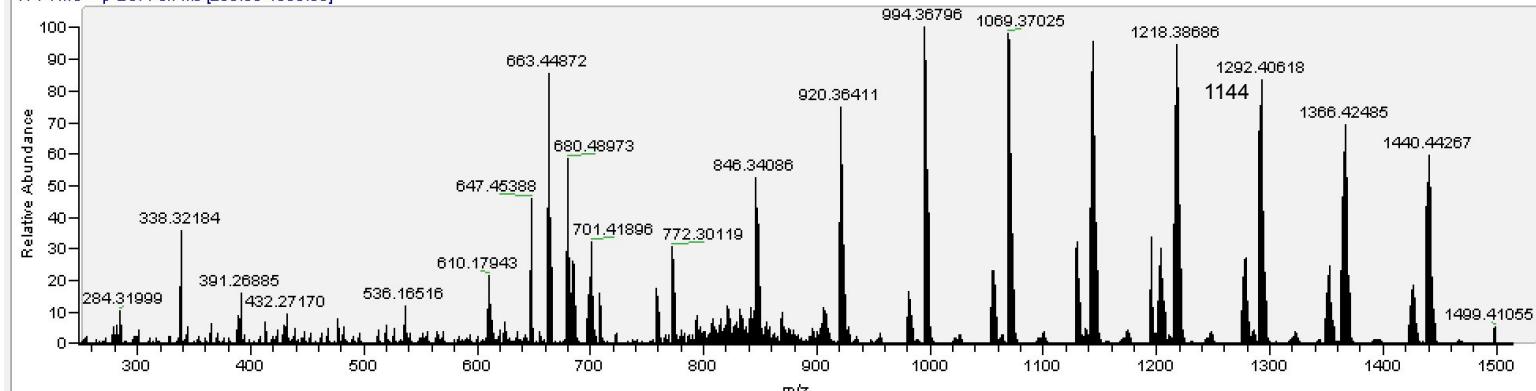
D:\MCF_Data\...\blank_IPA_bottle
Kinetex 2.6um_C18_3cmx2.1_IPA_MP_Formate

05/17/16 16:11:50

RT: 0.00 - 29.00 SM: 7B



blank_IPA_bottle #6565-11084 RT: 18.17-26.23 AV: 424 NL: 3.88E6
T: FTMS + p ESI Full ms [250.00-1500.00]



Sample vials/ plates

- Glass vials with insert
- 96 well plates
 - Zone-free covers
 - **No glue covers!**
- Check LC injector regularly

Carryover

- Blank injection after QC's
- Optimize needle wash
 - Use solvent mix
 - Compatible with sample type
 - Methanol:ACN:IPA:water
- reconstitution solvent compatibility with mobile phase

ESI source



- Lower pH
- Lower buffer conc
- Sample clean-up
 - High solvent
 - SPE
- Column bleeding, change

Buffer 20 mM Ammonium acetate and acetonitrile

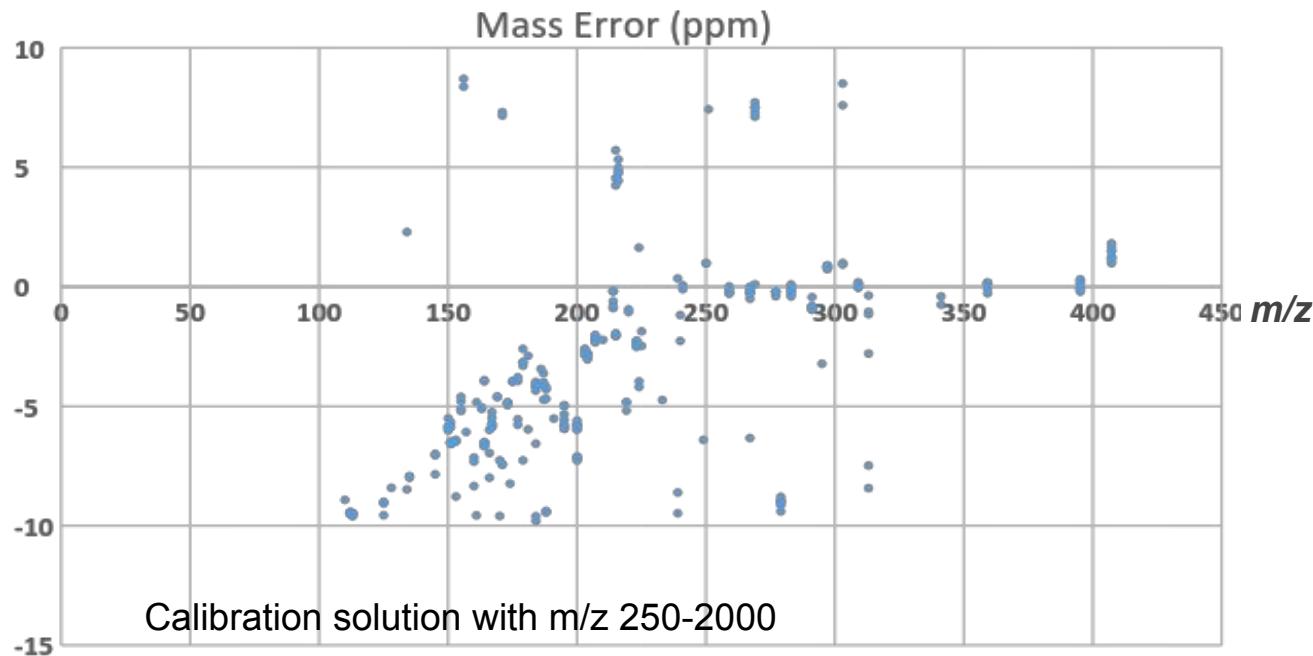
Clean-up

- Depending on samples analyzed
 - Clean MS source after every 300 injections
- Check conditions of your instrument
 - Track noise (background m/z)
 - Track internal standard variation
- Clean/flush LC system, injector
- Optimize ideal time for cleaning MS
- Follow by calibration and system suitability

Mass spec parameters

Calibration

- Calibrate the required mass range!
- Use custom calibration if necessary



Optimize MS parameters

- Flow-based parameters
 - Source temperature
 - Gases
 - Capillary temp
- MS scan related
 - m/z range (critical for orbitraps)
 - Ion transfer (AGC, injection time)
 - Scan speed vs resolution (critical for sensitivity)
 - May vary for positive and negative mode

MS response range

- Quantitation range/ dynamic range
- Metabolites in any sample has range of concentrations
- Optimize injection volume for sample type
 - Linear MS response
 - 8-12 uL for 10^6 cells or 10mg liver tissue
- Avoid sample overloading!
- Balance: sensitivity with detector saturation
- Optimize with standard mixture response

Different scans used to acquire MS/MS data

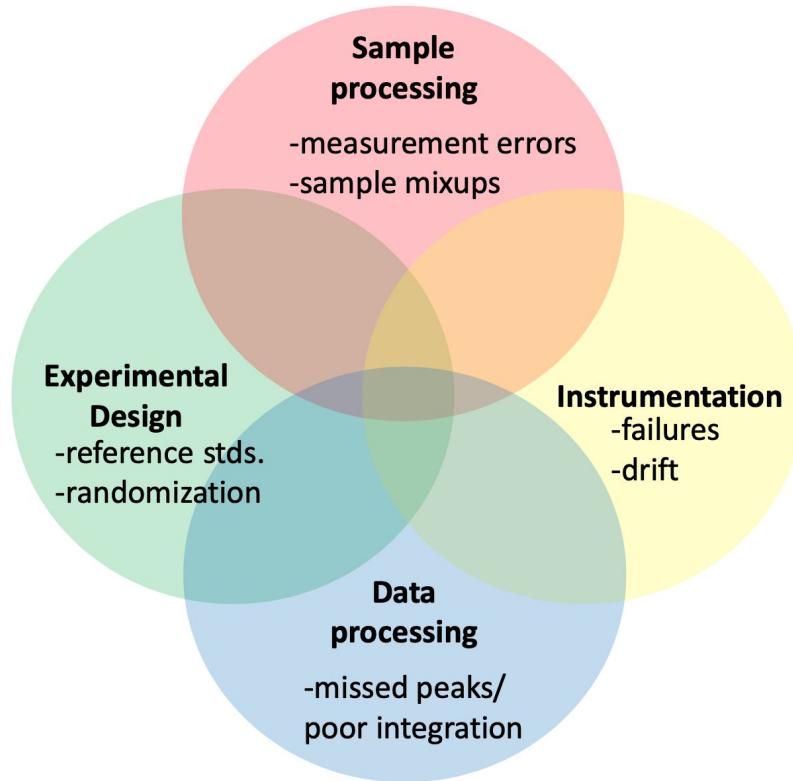
- Full scan
- DIA
 - No precursor selection
 - Top 5 or 10 most intense MS/MS
- DDA
 - Known precursor
 - PRM on QE

Not required to acquire MS/MS data for every sample

- Use pooled QC samples
- Generate candidates from MS data
- Use targeted inclusion list for quality MS/MS

QC system

**Quality control over
entire workflow**



Source: NIH website

Quality Control (QC) overview

QC system 1: Injecting QCs after each 6 injections (20% injection of total analysis sequence)

QC system 2: Internal standard spiking:

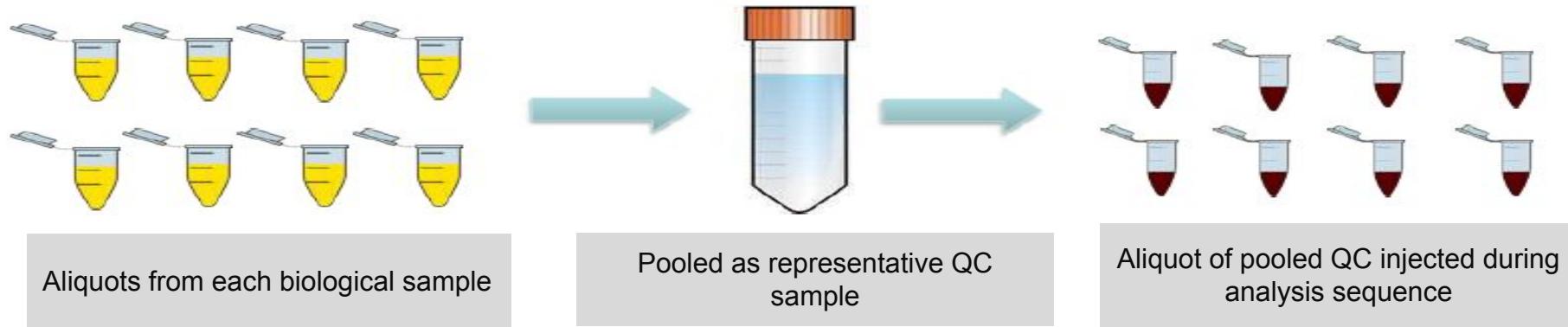
- a) sample prep → track sample prep variations
- b) Reconstitution → track instrument injection variations

QC system 3: Replicates (n=2, 3 or 5 technical replicates)



Acceptance of analysis sequence/ data
(PCA plot and % CV evaluation)

QC System 1: pooled QC

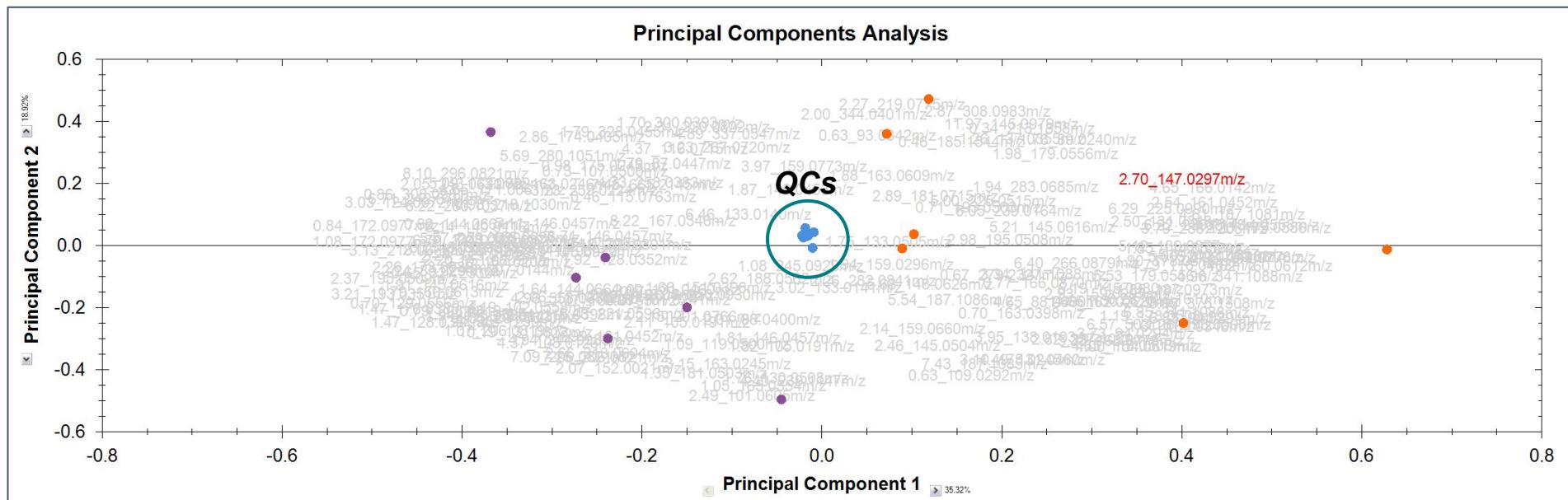


Other Options:

- Artificial QC from mixture of standards and available matrix
- Commercial serum/ plasma

QC System 1: pooled QC

Representative PCA plot of 6 pooled QCs after every 6 samples



Data from metabolomic analysis of 35 suspension cell extract samples sequenced with 6 pooled QCs over 48 hrs of total runtime

QC System 2: Internal standards (IS) spiked in biological samples to track sample extraction, recovery and analysis variations

- Glycolytic metabolites, Amino acids (HILIC-LC-MS)

- - ¹³C Creatinine and 13C pyruvate
 - Uridine-13C9,15N2 5'-triphosphate
 - ¹³C,15N Cell free amino acid mixture

- Lipids (RP lipid + ve mode)

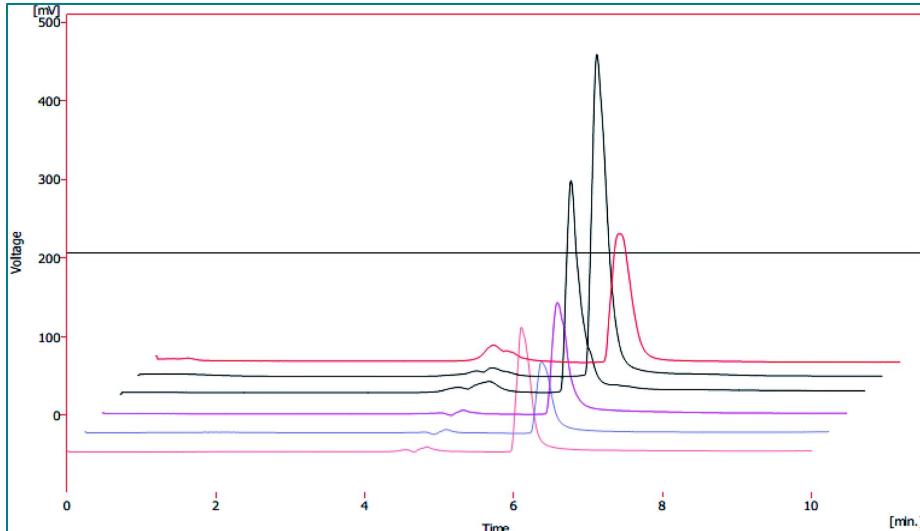
- - ¹³C labeled triolein
 - d5-DG ISTD Mix I (Avanti LM-6001)
 - Cardiolipin Mix I (Avanti LM-6003)
 - d5-TG ISTD Mix I (Avanti LM-6000)
 - PIP2 (37:4)
 - Fatty acids (RP lipid –ve mode)
 - ¹³C Palmitic acid

- - Steroids, Purine, drugs (RP + ve mode)

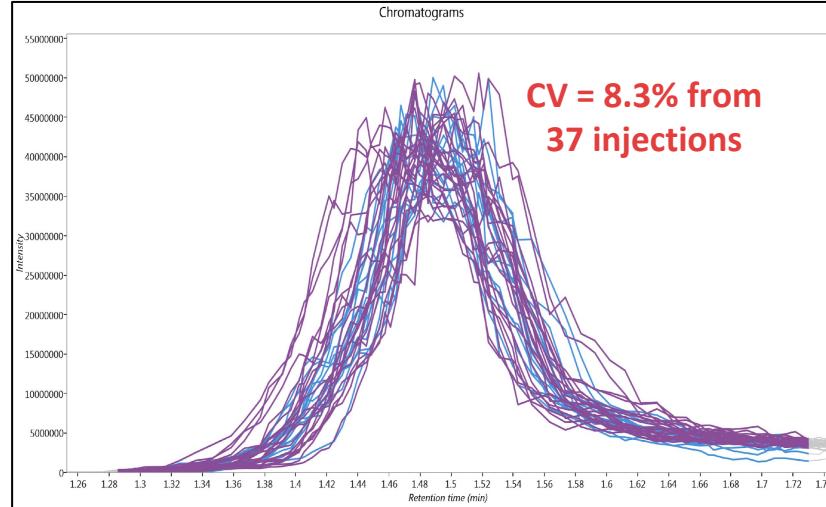
- ¹³C Testosterone
- Sulfachloropyridazine
- Sulfadimethoxine
- Amitriptyline

Quick QC check

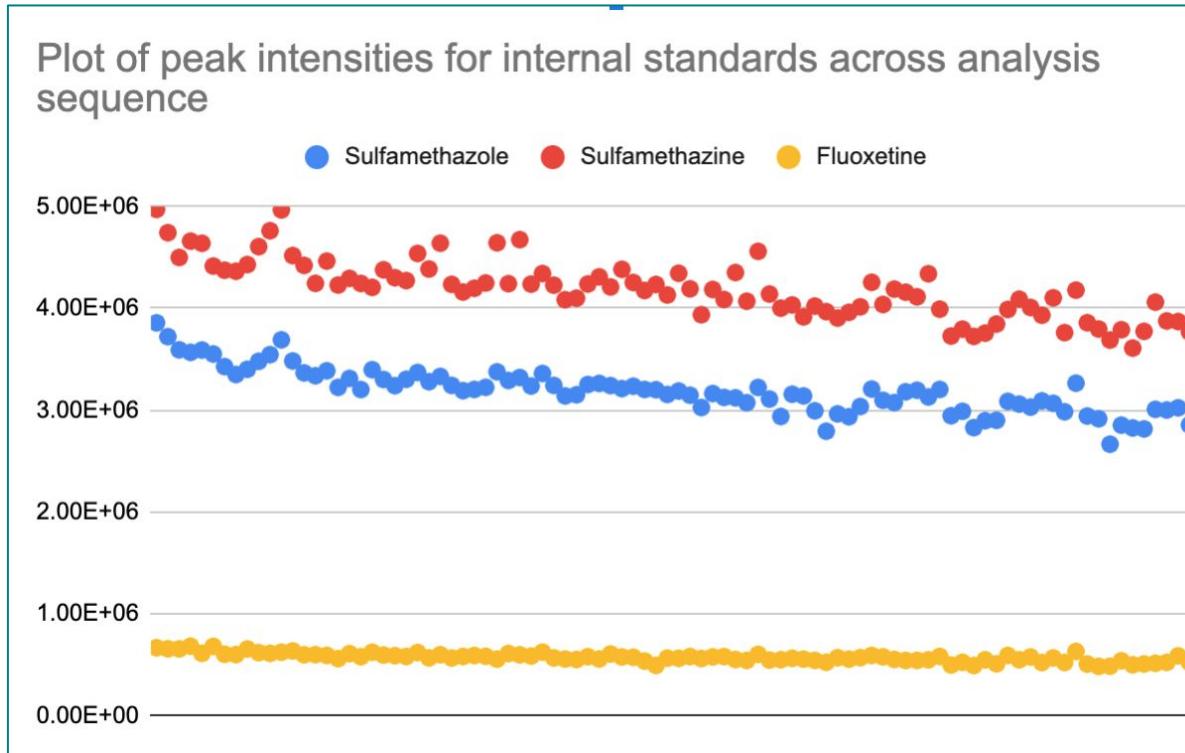
- Visual inspection → unusual patterns
- Chromatogram overlay TICs
- Check EICs for IS and knowns



¹³C Creatinine spiked IS variation in all Samples and QC



Internal standard variation



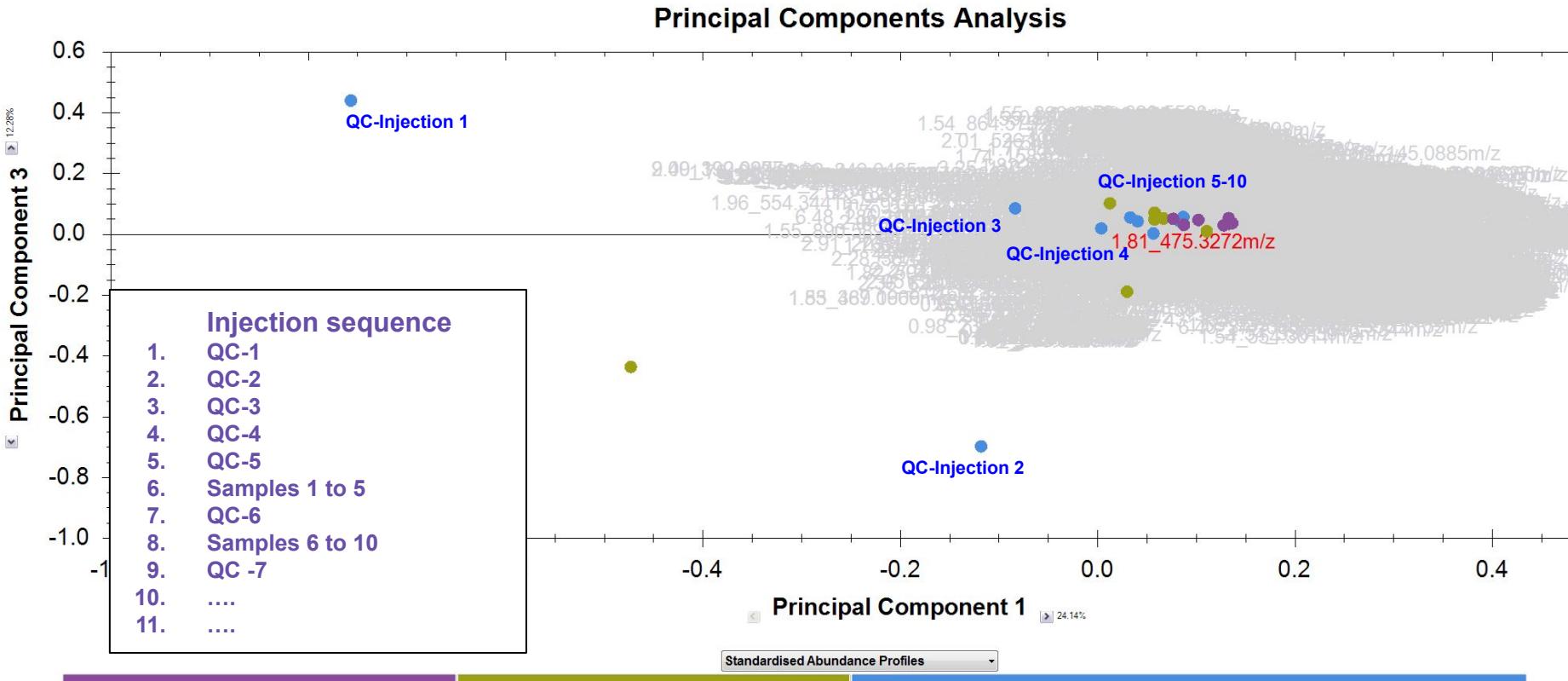
Multiple internal standards across chromatography and m/z

Data from 94 injections of bacterial cultures

QC System: acceptance criteria

- **Randomization:** Samples analysis sequence is randomized
- **System suitability:** Respective IS mixture (in previous slide) injected before analysis
- Blank and QC samples are 20-25 % of total batch
- **Prep/ Injection Replicates (n=3):** CV > 10-15 %
- **QC sample variation across analysis sequence:** CV > 20 % for targeted metabolites or at least 75 % features (for non-targeted analysis)
- **Spiked IS variation:** CV > 20 % for all spiked internal standards
- Data normalization if necessary using IS response or QC variation

PCA plot for QC injections



LC-MS analysis sequence

- Randomize analysis sequence
- LC-MS run sequence
 1. Solvent Blank (*check background*)
 2. System suitability (*standard mixture to evaluate instrument condition*)
 3. Standard mixtures (*for targeted analysis*)
 4. IS mixture (*for targeted and non-targeted analysis*)
 5. Test pooled QC samples 5 injections (*for stabilizing system*) → number of injections optimize for your system
 6. Processed Blank (*for checking carryover*)
 7. Samples (1 to 5)
 8. Pooled QC
 9. Samples (5 to 10)
 10. Pooled QC (*continues...*)
 11. System suitability (*end of analysis system evaluation*)
 12. Standard mixtures (*end of analysis system evaluation for targeted analysis*)
- Do not inject samples directly without stabilization
- Avoid blanks in between samples → de-stabilizes the system

Long term reference QC samples

- Sample available in quantity over a long period
 - Store aliquots
 - Prep and analyse during each sequence
- SRM1950 and NIST metabolomics ref material
 - Expensive
 - Suitable for interlaboratory comparisons
- Prepare long term ref QC suitable for your lab

Data analysis

Software

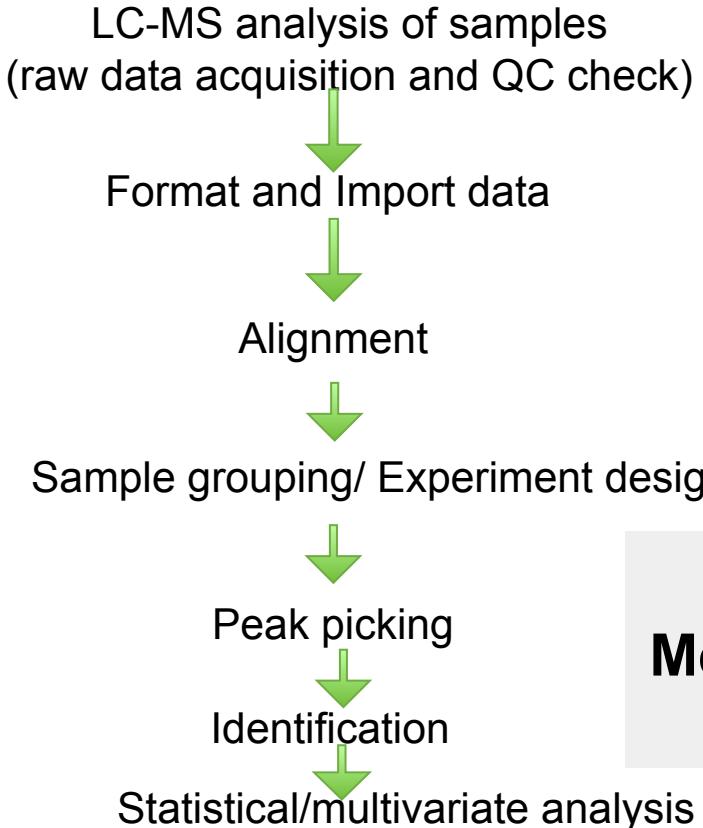
Commercial/ Freeware

- Progenesis QI (Untargeted)
- Xcalibur
- Thermo Compound Discoverer

Open-source

- Mz mine
- XCMS
- EI Maven
- Skyline (Targeted)

Data analysis: Non-targeted profiling



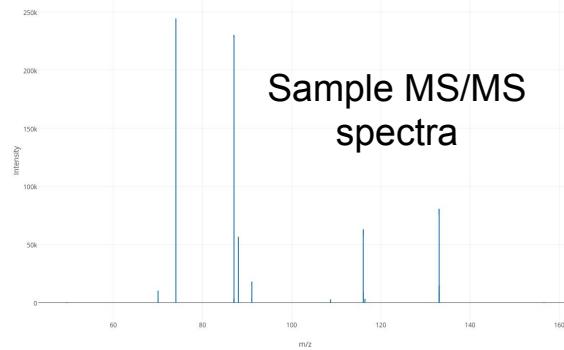
**Generate List of
Metabolite markers of
interest**

Data Normalization

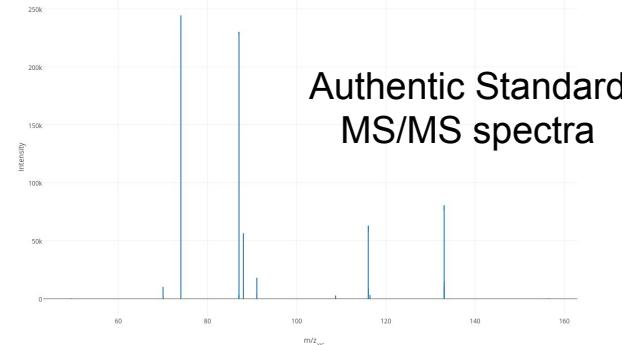
- Sample amount; cell numbers, OD, protein conc
- Internal standards
- Total intensity (TIC)
 - Better version; total compound/feature (less noisy)

Metabolite Identification

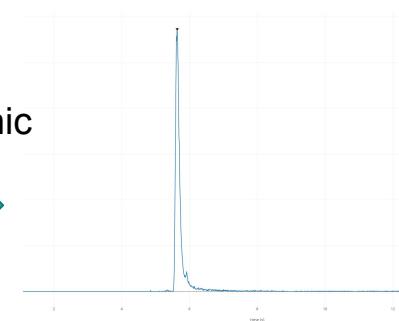
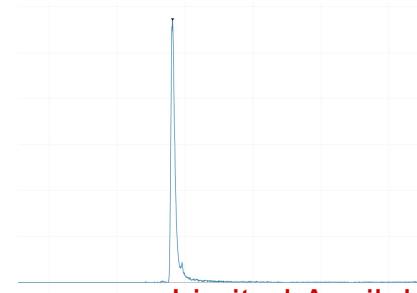
Level 1 ID



Spectral Matching

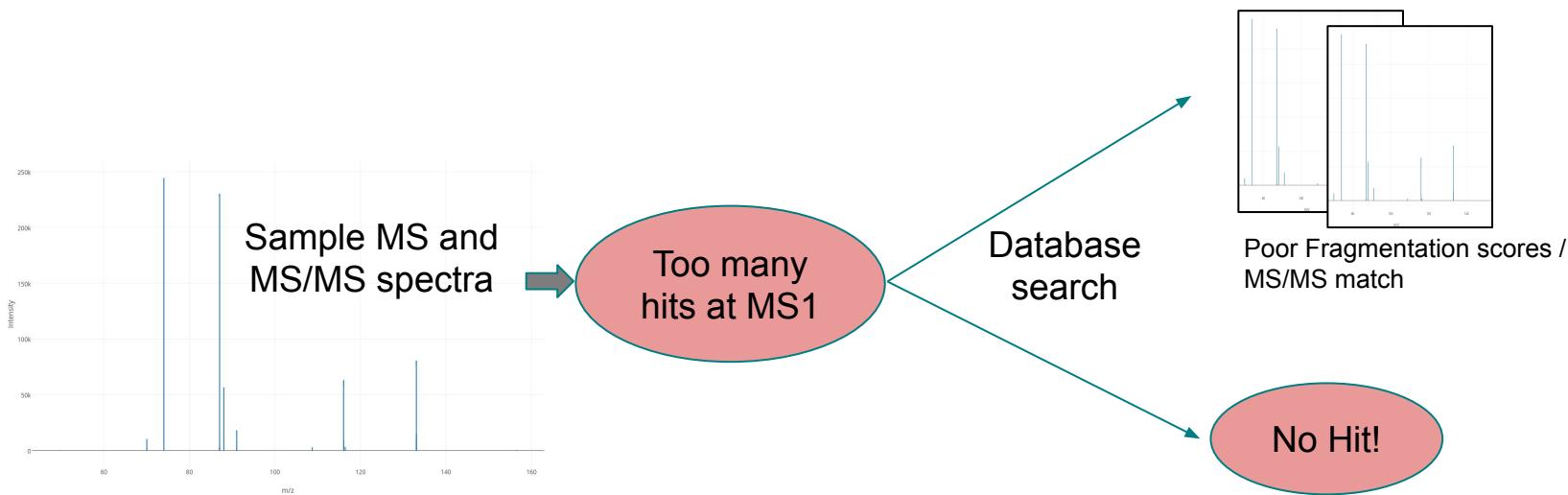


Matching chromatographic profiles



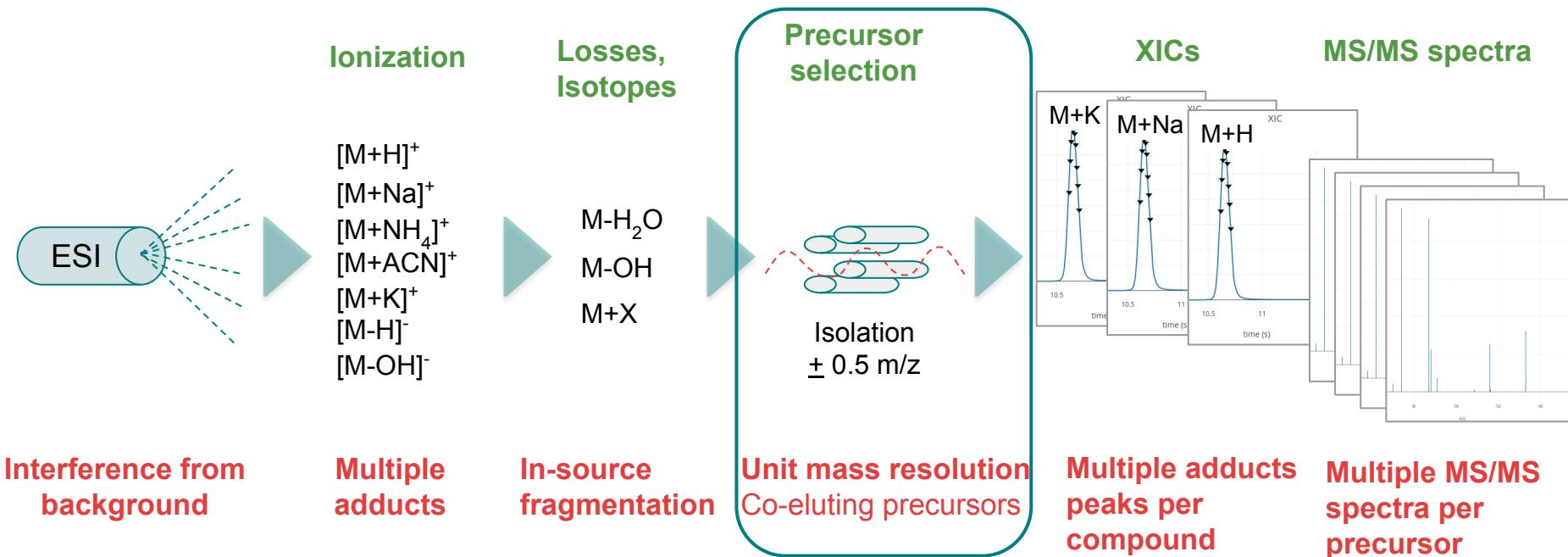
L.W. Sumner et al; Metabolomics (2007)

Spectral database search for Untargeted metabolomics

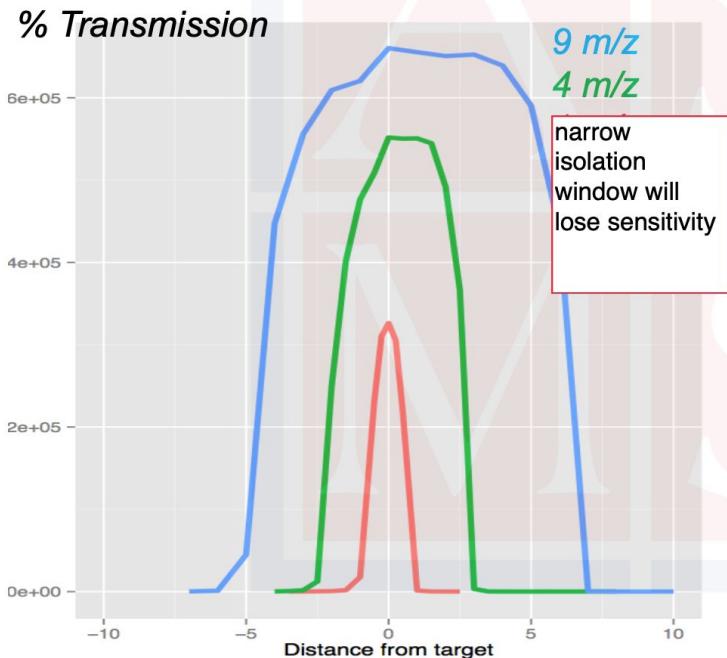


- Noisy features/ peaks
- Public libraries are not specific to Metabolomics study of your interest
- **Non-transferable LC-MS/MS spectra: different LC-MS conditions**
- False positive hits for exogenous compounds

Challenges of Metabolite ID in LC-MS/MS data



What causes “contaminated” metabolomic MS/MS spectra?

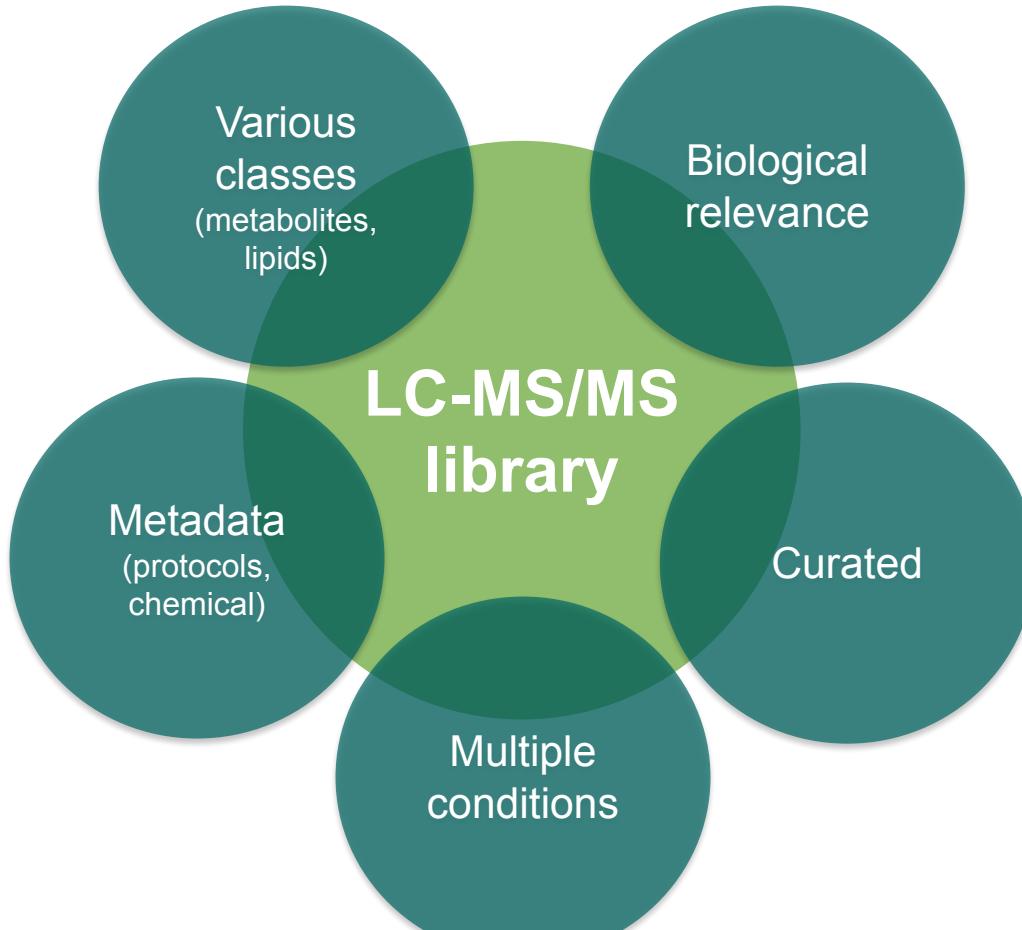


Co-eluting ions that have m/z values which are both isolated in the collision cell for fragmentation.

Adapted From ASMS 2018 Workshop by Gary Patti

Curated open-access LC-Orbitrap-MS/MS spectral library

LC-MS/MS spectral library for metabolomics



Available online ESI-MS/MS spectral libraries

Library	Compounds with spectra	Total MS/MS spectra	Orbitrap	LC info	Open access
mzCloud	6,585	2,045,858			
NIST14	9,345	234,284			
Metlin	14,000	72268			
Massbank	15,828	43,734			
GNPS	10,587	18,263			
HMDB	3,500	7228			

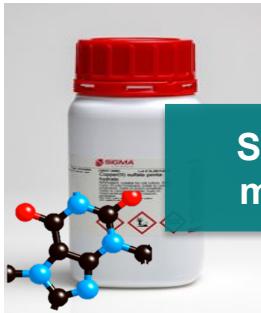
No fully open-access LC-Orbitrap-MS/MS library
for endogenous metabolites

Yes

Limited

No

Creating a spectral library: Our workflow



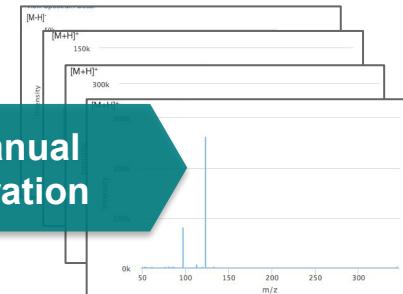
Standard mixtures



LC-MS method development



Analysis of standard mixtures



Manual curation

800 standards
10-15 in a mix

8 LC-MS protocols
ESI +ve, -ve, HILIC, C18

Evaluate XICs
Curate MS/MS spectra

20,000 €

3 months of development

300 h analysis time

10 min per spectrum

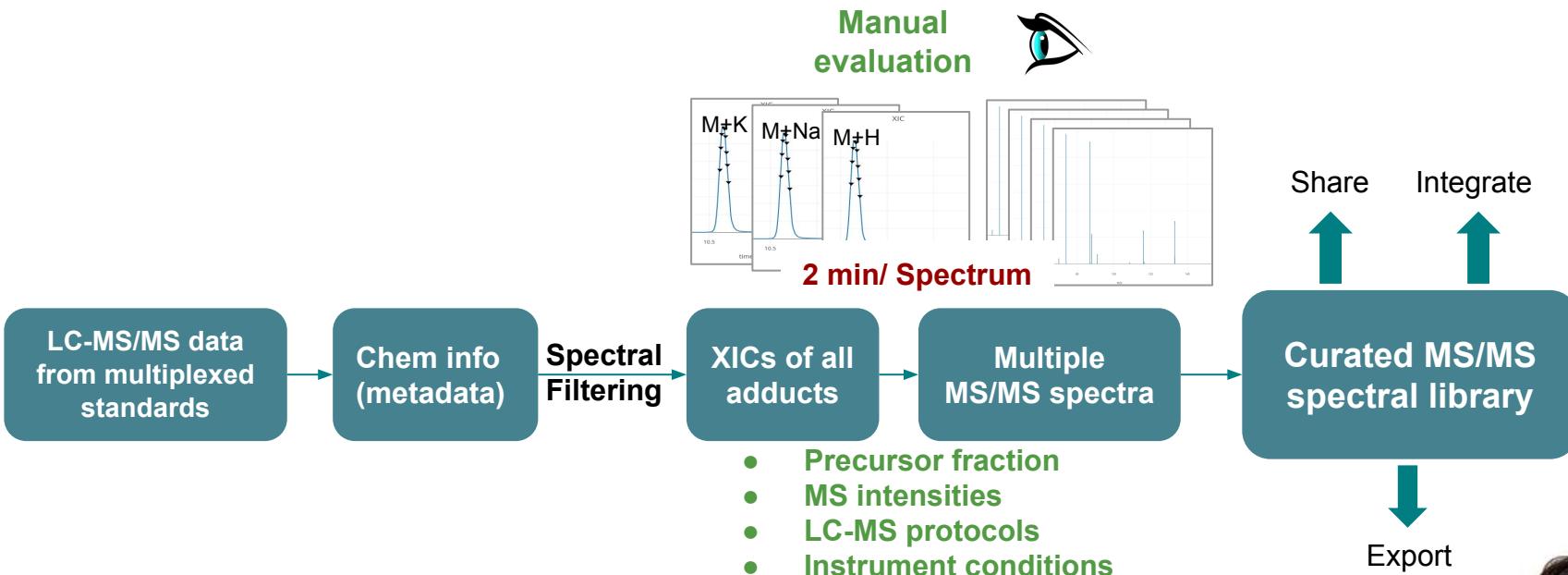
Resources

Time

Efforts

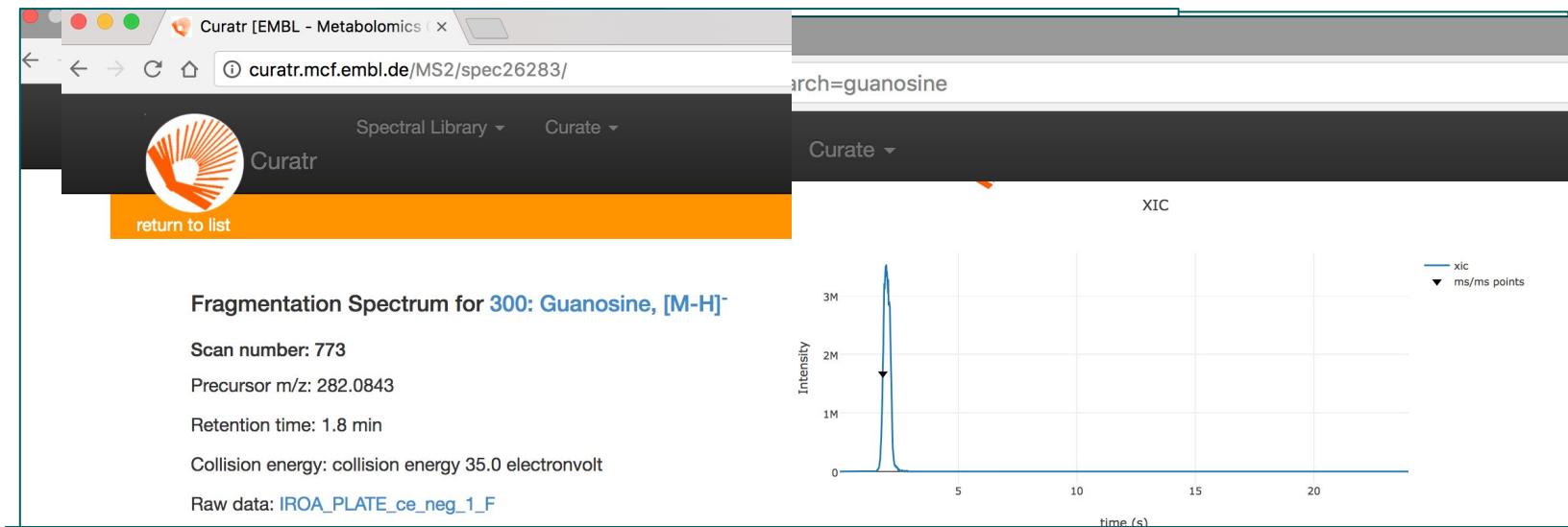
Our solution: web application Curatr

<https://github.com/alexandrovteam/curatr>



Andrew
Palmer

Web interface of Curatr



No Additional Software
Curate online
Easy Visualization
Fast

Date added: April 8, 2016

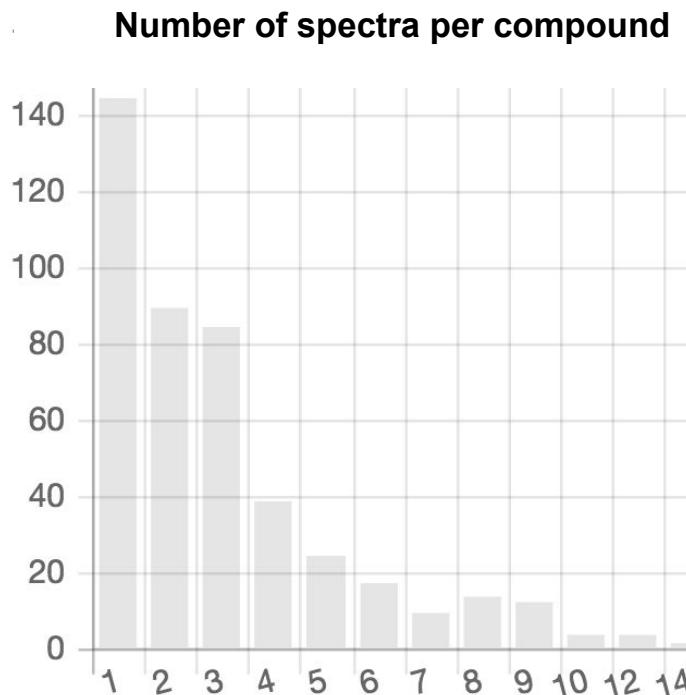
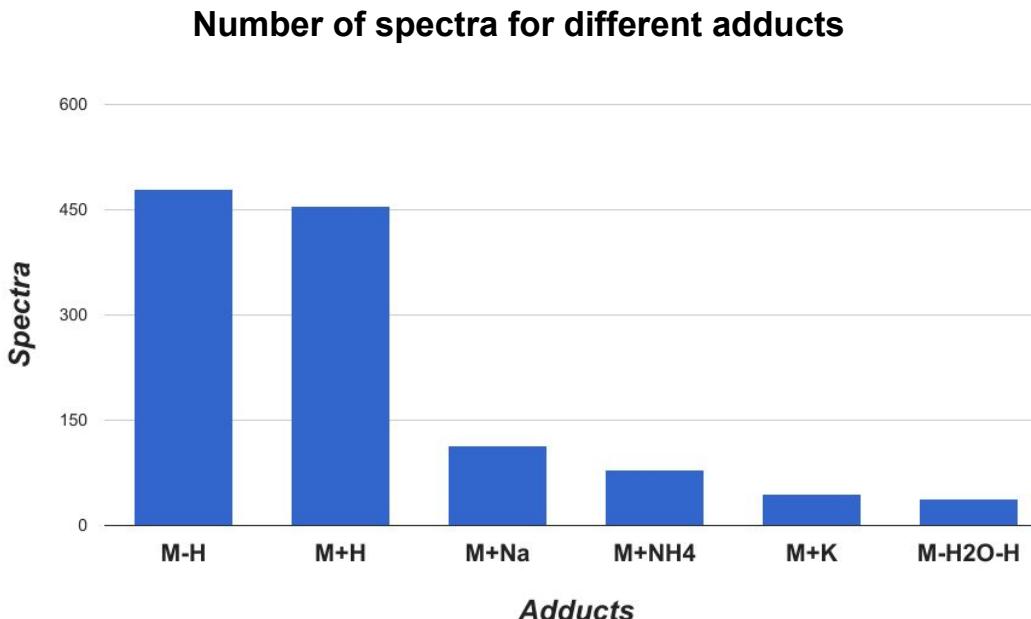
Date curated: April 10, 2016

Curator: prasad

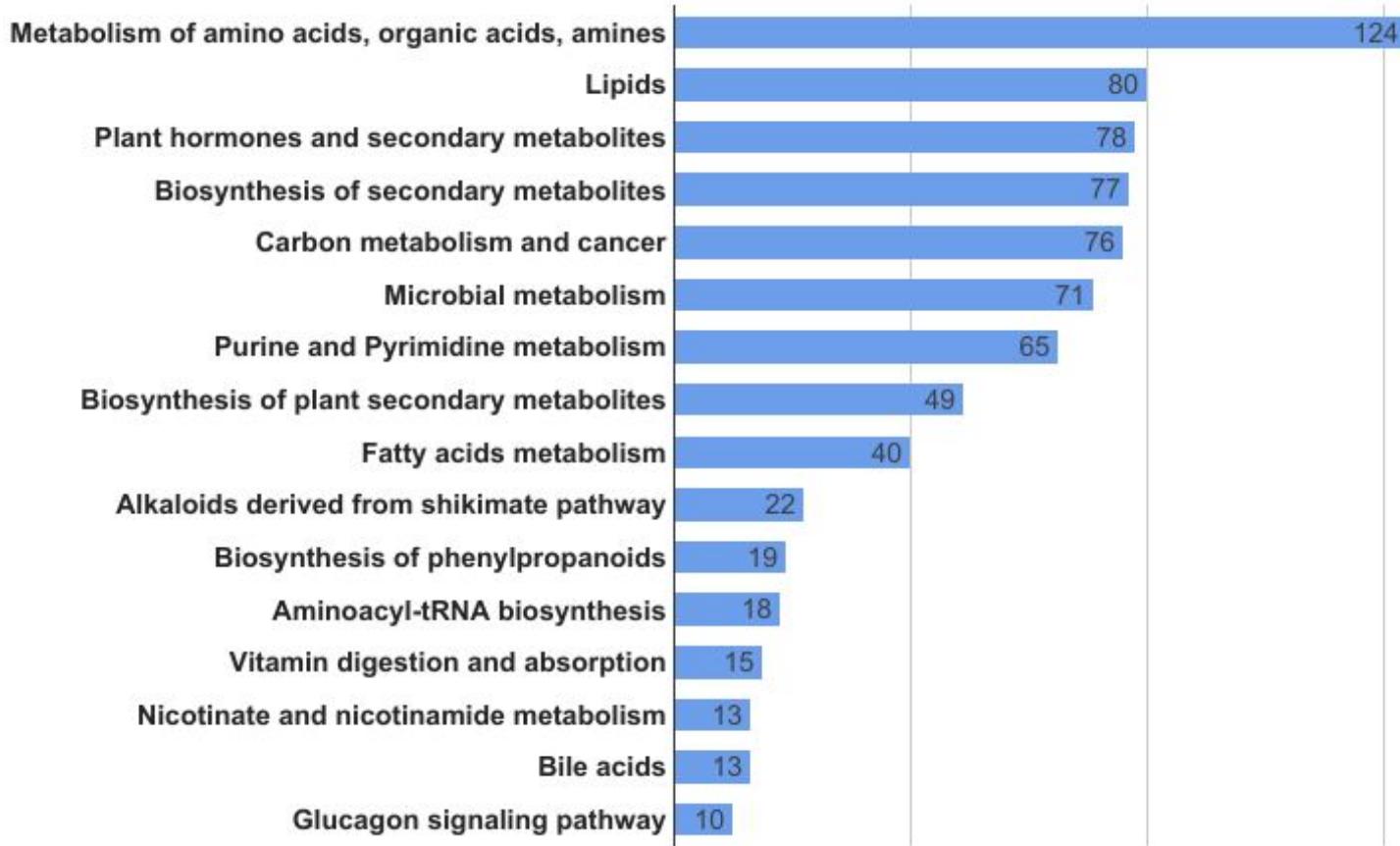


EMBL-MCF spectral library: Statistics

1,543 spectra for 439 compounds



Covered KEGG pathways

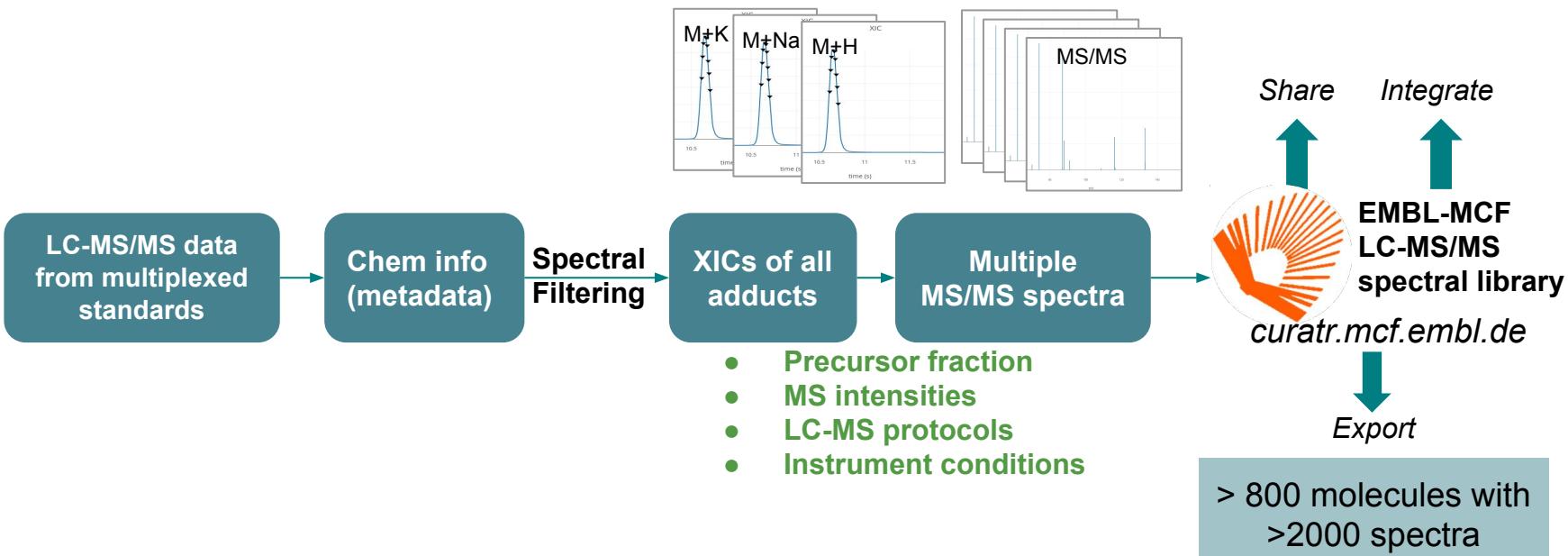


Open access, export and sharing

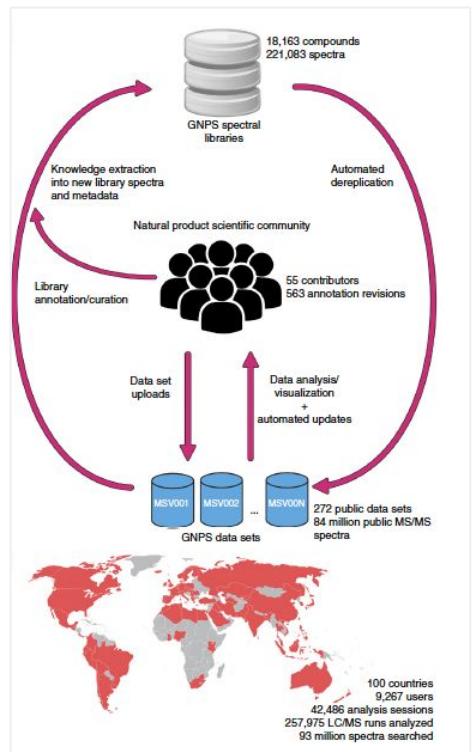
<http://curatr.mcf.embl.de>

- Open access
 - Creative Commons license (CC-BY)
- Downloadable in various formats
 - .TSV (as spreadsheet)
 - .MGF (spectral search)
 - **.MSP** (Text-based NIST)
- Integrated with spectral repositories
 - GNPS: included
 - MetaboLights: export
 - MassBank EU: on-going

EMBL-MCF LC-MS/MS spectral library



Integration with ‘Global Natural Product Social Molecular Networking (GNPS)’



Wang et al. *Nature Biotechnology* (2016)

Molecular Networking: A Graph-Based Tool to Explore Spectral Similarity in Liquid Chromatography–Tandem Mass Spectrometry (LC-MS/MS) Data from Molecular Mixtures

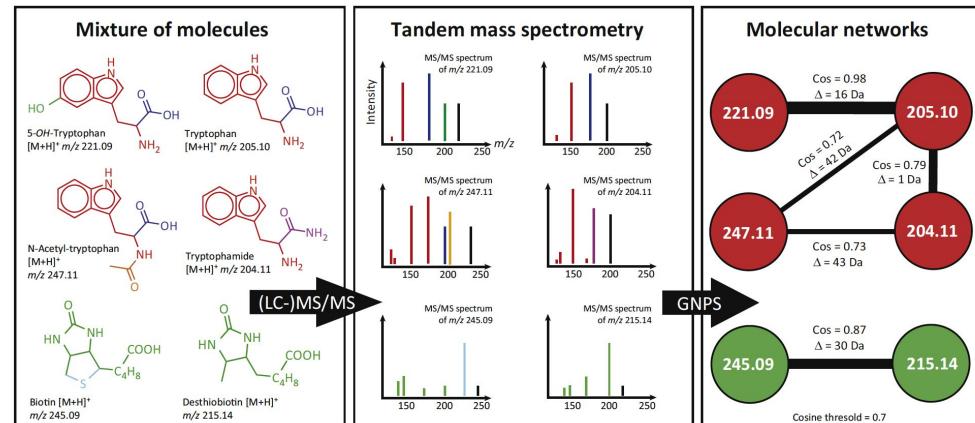


Figure 1. Molecular networks from LC-MS/MS spectra of tryptophan and biotin derivatives (from EMBL MCF spectral library on GNPS). The interactive view of the molecular network can be visualized directly on GNPS via the following link <http://gnps.ucsd.edu/ProteoSAFe/status.jsp?task=0415e5eef1d68149297b8aef26a480af9>.

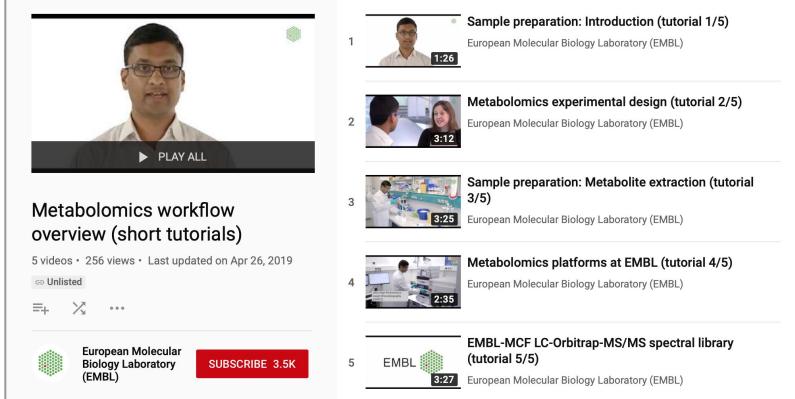
Quinn et al. *CellPress-Trends in Pharmacological Sciences* (2016)

UC San Diego

Resources

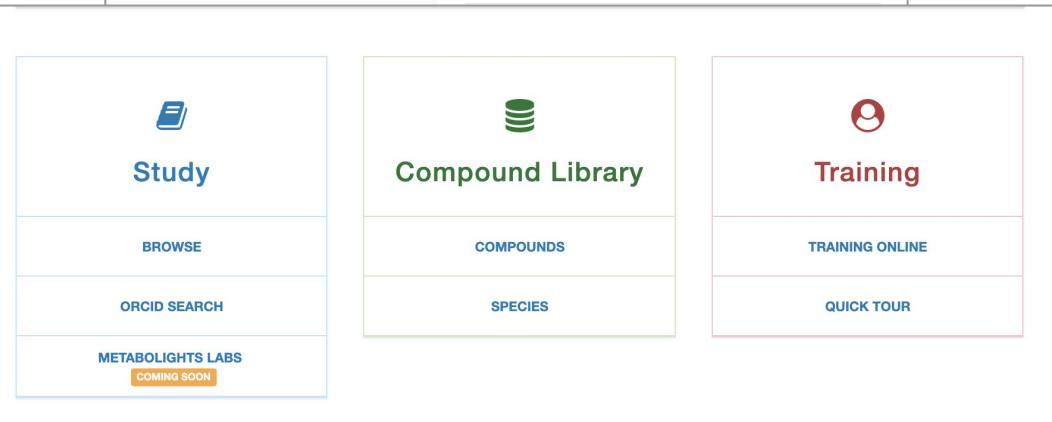
E-learning protocols
On [YouTube](#) EMBL channel →

- EMBL-EBI MetaboLights repository
- Metabolomics-forum.com
- Metaboanalyst



The screenshot shows a YouTube channel page for 'Metabolomics workflow overview (short tutorials)'. It features a video thumbnail of a man speaking, a 'PLAY ALL' button, and a list of five videos:

- 1. Sample preparation: Introduction (tutorial 1/5) by European Molecular Biology Laboratory (EMBL), 1:26
- 2. Metabolomics experimental design (tutorial 2/5) by European Molecular Biology Laboratory (EMBL), 3:12
- 3. Sample preparation: Metabolite extraction (tutorial 3/5) by European Molecular Biology Laboratory (EMBL), 3:25
- 4. Metabolomics platforms at EMBL (tutorial 4/5) by European Molecular Biology Laboratory (EMBL), 2:35
- 5. EMBL-MCF LC-Orbitrap-MS/MS spectral library (tutorial 5/5) by European Molecular Biology Laboratory (EMBL), 3:27



The MetaboLights website interface includes three main sections:

- Study**: Includes BROWSE, ORCID SEARCH, and METABOLIGHTS LABS (COMING SOON).
- Compound Library**: Includes COMPOUNDS and SPECIES.
- Training**: Includes TRAINING ONLINE and QUICK TOUR.

Acknowledgement

Alexandrov Team/ MCF

Theodore Alexandrov

Andrew Palmer

Maria Naumenko

MCF users

Dorrestein Lab - UCSD

Past group members

Thank you

'Heidelberg Molecular Life Sciences (HMLS) Investigator Award' for Advion Nanomate instrument

NIH Common Fund Metabolomics program for free ¹³C standards

EMBL Core Facilities

