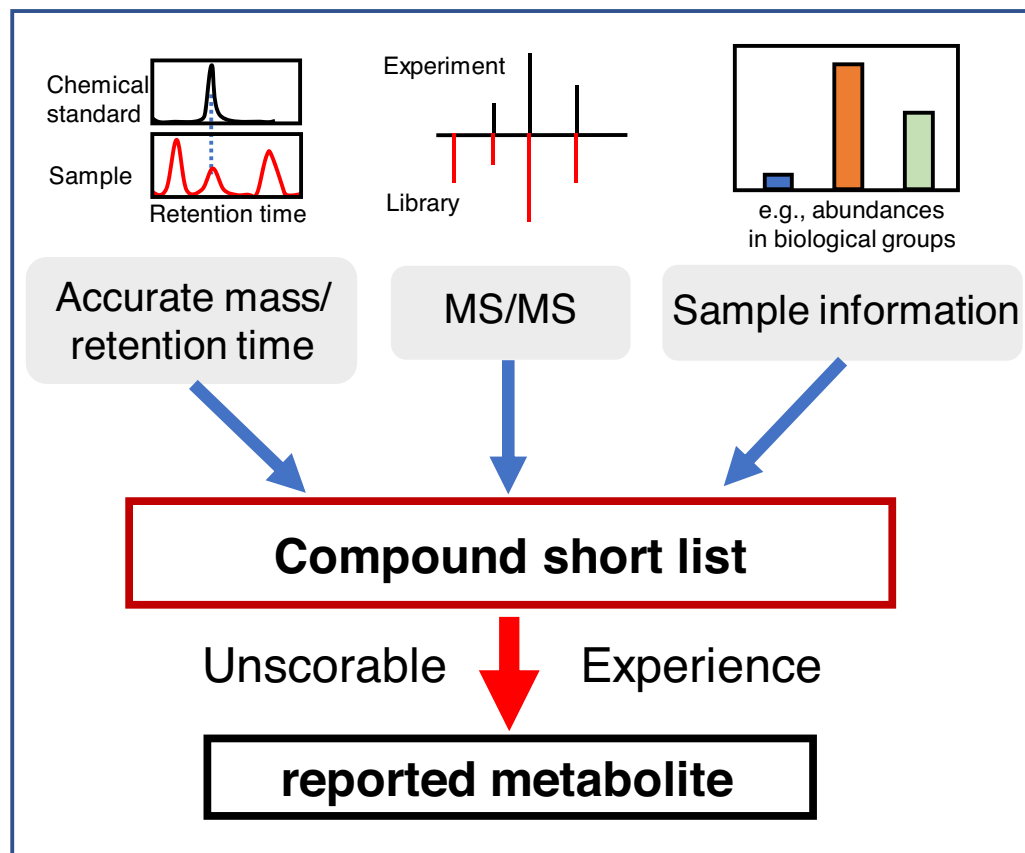


Metabolomics annotation and metabolite reporting

- Annotation process
- Computational annotation
- Reporting standard

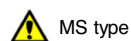
Challenges, progress and promises of metabolite annotation for LC-MS-based metabolomics

Romanas Chaleckis^{1,2}, Isabel Meister^{1,2}, Pei Zhang^{1,2} and Craig E Wheelock^{1,2}



feature

0% ? ? ? 100%
Identification confidence



MS type



Calibration

(a) Accurate mass

< 5 ppm from theoretical mass

no

Check

- ppm error in the dataset (\downarrow m/z (<100) tends to \uparrow ppm)
- ppm error for IS and well known metabolites

yes



LC method



Sample type

(b) Retention time

< 0.1 min RP, < 0.3 min HILIC from chemical standard

no

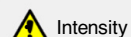
Check

- RT shift of IS & common metabolites compared to chemical standard
- Order of closely eluting compounds with characterized chemical standards
- Coeluting isobaric compounds require further confirmation (e.g., MS/MS and/or sample information)
- Fragment of another compound (e.g., adenine-adenosine)

no closely or co-eluting compounds with same m/z

MS/MS available

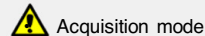
yes



Intensity



Collision energy



Acquisition mode

Several major fragments matching

no

Check

- database and algorithm used for MS/MS search
- co-eluting compounds complicating the spectra

yes

(d) Sample information



Not very reactive or unstable?

(aldehydes and redox compounds (e.g., containing -SH group) usually require dedicated protocols)



Likely to ionize at the applied LC-MS conditions?

(organic acids are not likely to be detected in positive ionization mode; polyamines and carnitines in negative mode)



Expected concentration in the sample >100 nM?

(low nM concentrations are usually not detected by untargeted metabolomics due to methods sensitivity)



Not present in the sample blanks, reasonable S/N ratio, low variation in the QC samples?

(presence in the blank samples and high variation hints to contaminant rather than metabolite from the sample)



Input from metabolic network analysis, isotope tracing, other data (gene, protein) etc.

metabolite

0% 100%
Identification confidence

Computational annotation

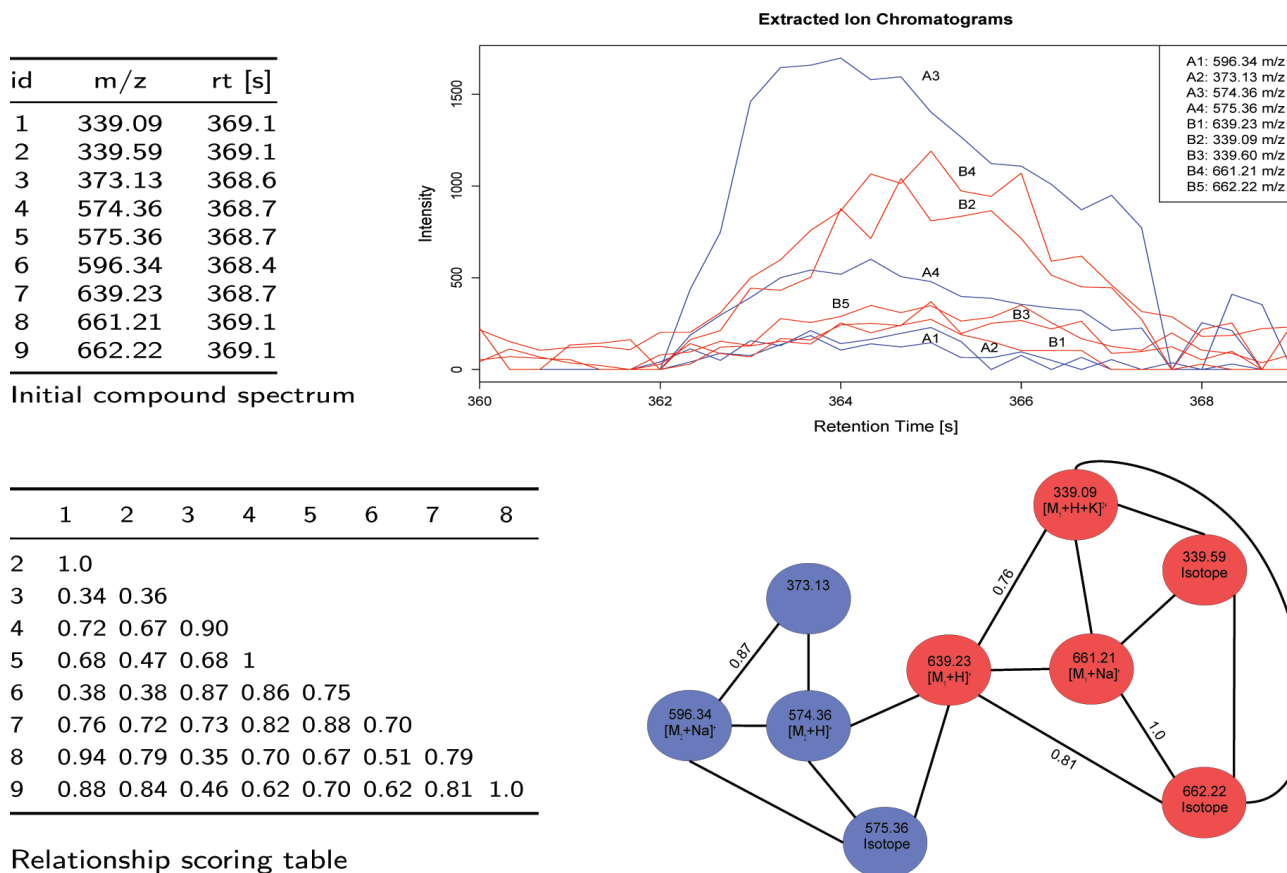
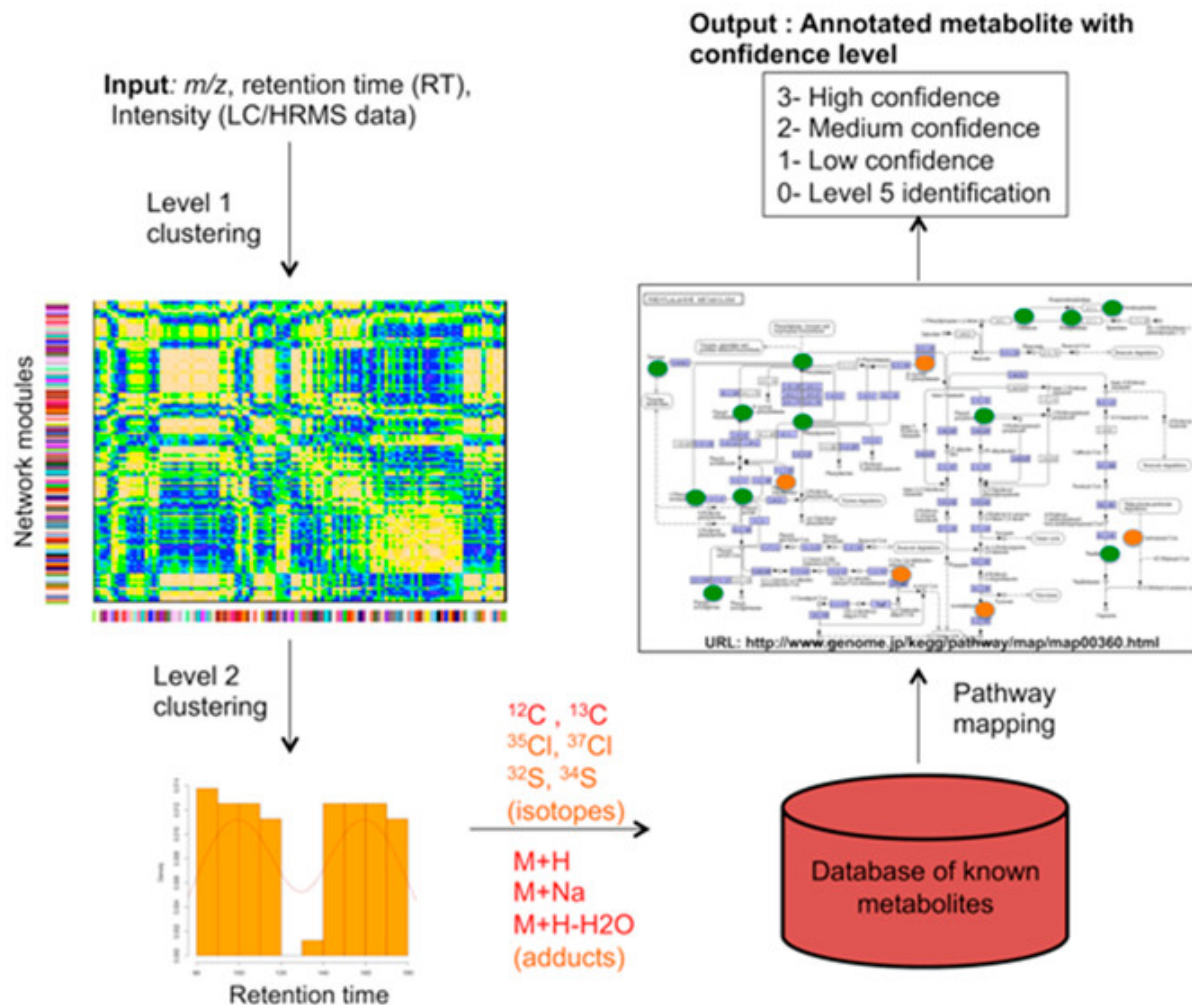
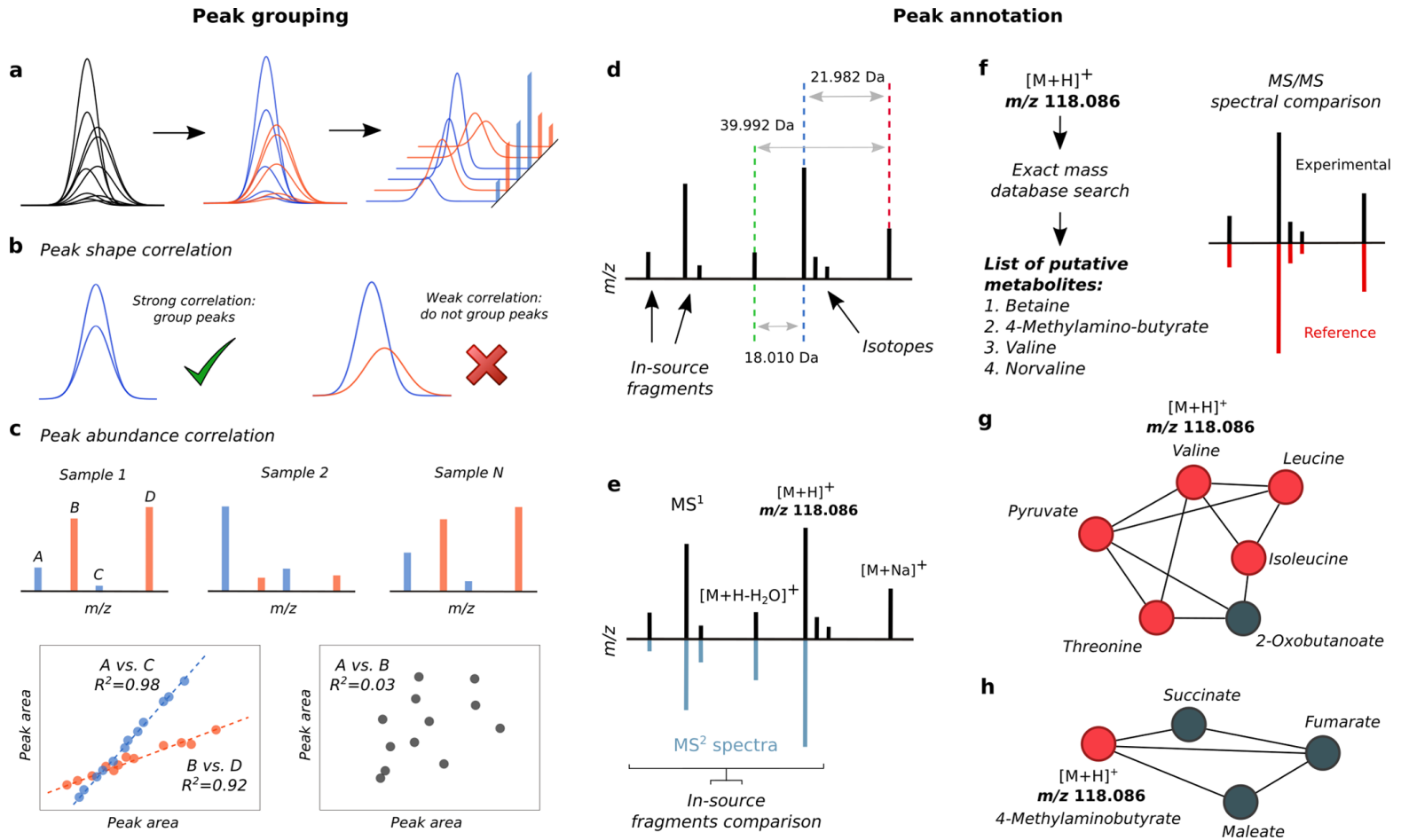


Figure 2. Schematic clustering of low-intensity features initially grouped by retention time into a single compound spectrum. Top left: the features, initially grouped by retention time. Top right: the EICs of all features. The labels A and B correspond to the result after graph clustering. Bottom left: the scoring matrix, used as edge weights in the graph. Bottom right: the relationship graph, where edges indicate an above-threshold score. The node labels include the ion species annotation, and the node color shows the graph separation after refinement with the LPC algorithm (A = blue, B = red).

Computational annotation



Computational annotation



Sumner et al 2007

1. Identified compounds. A minimum of two independent and orthogonal data relative to an authentic compound analyzed under identical experimental conditions are proposed as necessary to validate non-novel metabolite identifications
2. Putatively annotated compounds (e.g. without chemical reference standards, based upon physicochemical properties and/or spectral similarity with public/commercial spectral libraries).
3. Putatively characterized compound classes (e.g. based upon characteristic physicochemical properties of a chemical class of compounds, or by spectral similarity to known compounds of a chemical class).
4. Unknown compounds—although unidentified or unclassified these metabolites can still be differentiated and quantified based upon spectral data.

Sumner, L.W., Amberg, A., Barrett, D., Beale, M.H., Beger, R., Daykin, C.A.: Proposed minimum reporting standards for chemical analysis Chemical Analysis Working Group (CAWG) Metabolomics Standards Initiative (MSI). *Metabolomics* **3**, 211–221 (2007)

Schymanski et al 2014

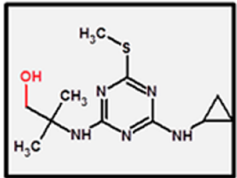
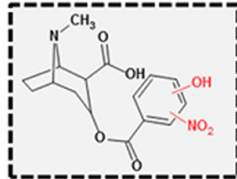
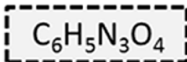
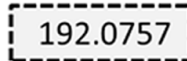
Example	Identification confidence	Minimum data requirements
	Level 1: Confirmed structure by reference standard	MS, MS ² , RT, Reference Std.
	Level 2: Probable structure a) by library spectrum match b) by diagnostic evidence	MS, MS ² , Library MS ² MS, MS ² , Exp. data
	Level 3: Tentative candidate(s) structure, substituent, class	MS, MS ² , Exp. data
	Level 4: Unequivocal molecular formula	MS isotope/adduct
	Level 5: Exact mass of interest	MS

Figure 1. Proposed identification confidence levels in high resolution mass spectrometric analysis. Note: MS² is intended to also represent any form of MS fragmentation (e.g., MS^e, MSⁿ).

Schrimpe-Rutledge et al, 2016

