Differential analysis for quantitative metabolomics using isotope-labeling and LC-HRMS: data processing strategies

Michel Wagner[1], Yasmin Boukhedimi[1], Leanne Ohlund[1], Tze Chieh Shiao[1], Amelie Vezina[1], Borhane Annabi[1], Alex P. Parker[2], Sarah Jenna[1], Rene Roy[1], Lekha Sleno[1]

[1]Université du Québec à Montréal, Montreal, Canada. [2]University of Montreal, Montreal, Canada

Introduction

Typical LC-MS-based metabolomics relies on untargeted workflows with chemometrics-based data processing:

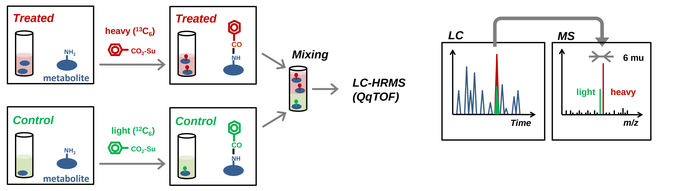
- Identify features (retention time – *m/z* pairs) discriminating sample groups (e.g. diseased *vs.* healthy)
- Limitations: no quantitative information and no structural information about each feature (*i.e.* identification)

We propose an alternative strategy by isotope-labeling:

- Derivatization of amines with benzyl-oxysuccinimide
- Two versions of each reagent, differing by 6 amu : light (only ^{12}C atoms) and heavy (containing six ^{13}C atoms).

Workflow for differential analysis:

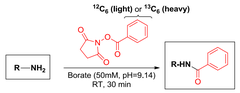
- Two different "states" of samples are extracted and labeled with a different version of the reagent
- Samples are pooled and analyzed by LC-HRMS using a QqTOF MS instrument



Methods

1. Labeling of amine metabolites

With benzyl-oxysuccinimide in borate buffer at room temperature.



2. Biological samples: HL60 cell cultures

Extraction of metabolites: with cold methanol and sonication

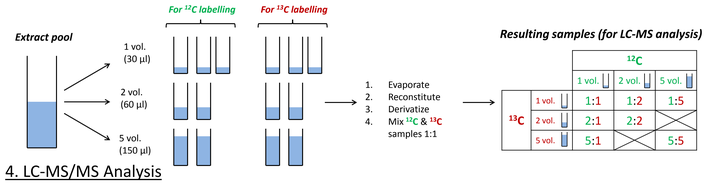
3. Relative quantitation

a) Extract pools

Sample extracts (each from 5 x 10^6 of cells) were pooled in 5 different tubes, defining 5 independent series.

b) "Dynamic range"

- Each pool extract was **aliquoted** in three different volumes (30, 60 or 150 µl, *i.e.* relatively 1, 2 or 5 vol.)
- Samples were **evaporated**, **reconstituted** in 120 ul of borate buffer/MeCN, 95/5 (v/v) and **derivatized** with either the light (^{12}C) or the heavy (^{13}C) labeling reagent.
- ^{12}C and ^{13}C-labeled samples were **mixed** in different combination, according to the initial extract volume.

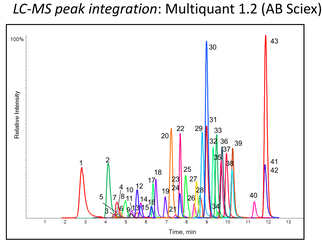


4. LC-MS/MS Analysis

HPLC system: Shimadzu Nexera. *LC column*: Cogent C18 (150 x 2mm, 4 µm)
Mobile phases: A: H2O + 0.1% HCOOH, B: MeCN + 0.1% HCOOH *Flow rate*: 0.4 ml/min.
Gradient: 5%B (0-2min), 25% (6min); 65%B (17min); 90%B (18 min).
MS instrument: AB Sciex 5600 TripleTOF.
Source conditions: ESI, positive ion mode. TEM=500°C. ISVF=5000V. GS1=GS2=50. CUR=35.
MS experiments: 1:TOF-MS, *m/z* 110-1000 2-6:MS/MS information-dependent acquisition, CE=30±10, DP=60

5. Data processing

Peak picking and pair finding: MsXelerator (MsMetrix) or MetabolitePilot custom tool (AB Sciex)
Reviewing of LC-MS data: PeakView 1.2 (AB Sciex)
LC-MS peak integration: Multiquant 1.2 (AB Sciex)



LC-MS known derivatized metabolites

	Metabolite
1	Histidine (His)
2	Arginine (Arg)
3	Lysine (Lys)
4	Cystine
5	Asparagine (Asn)
6	Serine (Ser)
7	Hydroxyproline (Pro-OH)
8	Glutamine (Gln)
9	Aspartic acid (Asp)
10	Glycine (Gly)
11	Oxidized GSH (GSSG)
12	Citrulline (Cit)
13	Methionine Oxidized (Met-O)
14	Glutamic acid (Glu)
15	Threonine (Thr)
16	Alanine (Ala)
17	Glutathione (GSH)
18	GABA
19	Dihydroxyphenylalanine
20	Proline (Pro)
21	GSSG 2x
22	Hydroxy-tyrptophan (Trp-OH)

	Metabolite
23	a-aminobutyric acid (AIBA)
24	Tyrosine (Tyr)
25	Dopamine (Dopa)
26	Hydroxy kynerinine
27	Valine (Val)
28	Methionine (Met)
29	Serotonin (Sero)
31	Ornithine 2x (Orn 2x)
30	Tyramine
32	GSH 2x
33	Lysine (Lys) 2x
34	Cystine 2x
35	Kynurenine (Kyn)
36	Isoleucine (Ile)
37	Leucine (Leu)
38	Phenylalanine (Phe)
39	Tryptophan (Trp)
40	Hydroxy-kynurenine 2x (Kyn-OH 2x)
41	Cysteine (Cys) 2x
42	S-benzyl cysteine
43	Tryptamine

Results

A Isotopic pair finding : comparison of two software packages

1. Experimental

Pair finding:

Both software find pairs according to a user pre-defined ratio and an associate tolerance.

The same sample (5:5) was processed with both software, expecting therefore 1/1 ratios ($^{12}C/^{13}C$).

Software parameters:

Some differences exist due to different software designs.

- Identical parameters:

- Time and m/z ranges: encompassed the ranges
- Δ m/z : 6.02 0.01
- Divergences mostly linked to ratios and tolerances:

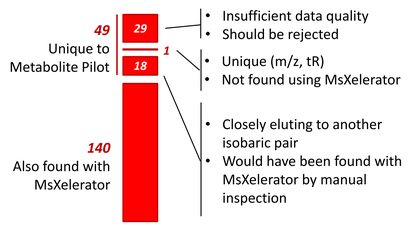
MetabolitePilot: based on peak height, ratios $^{12}C/^{13}C$ = 1/1 0.5 (i.e. 0.5-1.5)

MsXelerator: based on peak height and peak area ratios $^{12}C/^{13}C$ = 1/0.5 to 1/2

Data processing:

- **MetabolitePilot** required searches to be divided in several *m/z* intervals due to limitation in number of pairs that can be reported. Overall process relatively time consuming.
- **MsXelerator** required a preliminary step to import and convert raw data. One step data processing relatively fast.

2. MetabolitePilot : 189 feature pairs



3. MsXelerator : 484 feature pairs

- **Excellent reproducibility** for peak picking and pair finding was obtained between the five different series.

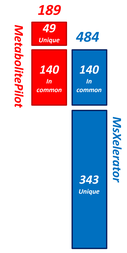
Sample	Pairs
5:5 (series a)	484
5:5 (series b)	512
5:5 (series c)	501
5:5 (series d)	499
5:5 (series e)	492
mean	497.6
CV	2.1%

- **Number of feature pairs** depend on their concentration in the sample

Sample	Pairs	% of 5-5 (series a)
2:2 (series a)	448	93%
1:1 (series a)	320	66%

4. Result summary

Number of isotope pairs obtained with both softwares and result overlap:



Given the larger number of features pairs reported by MsXelerator and that only one pair was missed, more thorough data processing was achieved with this software.

B Data inspection : results from MsXelerator

1. Data review in PeakView

List of feature pairs generated by MsXelerator was used to build a XIC list comprising feature name, RT and *m/z*. This list allowed reviewing LC-MS data in PeakView in order to remove irrelevant features and to add missed ones.

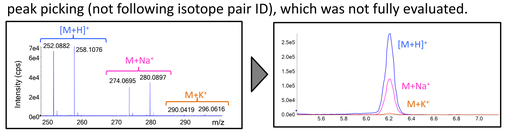
● Na+ / K+ adducts (102 features pairs removed)

a) Semi-automated search in Excel, search for ions:

- differing by 21.98195 (0.01) (Na^+) or 38.96371 (0.01) (K^+) & same RT

b) Verification in PeakView: by overlaying corresponding LC-MS traces.

Note: MsXelerator provides option to filter Na^+ and K^+ adducts during initial peak picking (not following isotope pair ID), which was not fully evaluated.



● Blank correction (80 feature pairs removed)

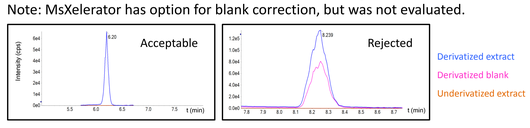
Comparison of peak heights of:

- Less concentrated sample extract
- Underivatized extracts (extract without labeling reagent)
- Reagent blank samples (all reagents except extract)

This was done for each of 5 series, each having both types of blanks.

Criteria: Pair was rejected if peak signal in either blank was >30% of derivatized extract. For each pair, at least 4 of the 5 series met the rejection criteria.

Note: MsXelerator has option for blank correction, but was not evaluated.



● Removal of features with low significance (9 pairs discarded)

Manual inspection in PeakView; related to peak with unacceptable S/N (≤ 3)

Results: - Relatively marginal (ca. 2% of the initial data)

2. Practical considerations

● Redundant data

In-source alteration of chemical structure of the analyte by several processes, such as **fragmentation** (e.g. loss of H_2O or HCO_2H) and **oxidation** (addition of O). Consequence: peak pairs not from intact labeled metabolites

Investigation (similar to Na^+/K^+ adducts):

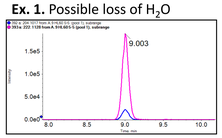
a) Semi-automated identification of possibly related ions in Excel, based on known mass differences

b) Verification of the results in Peakview

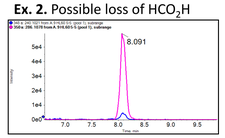
Results (the pairs identified were not discarded from the list):

Process	- H2O	- CO2	- HCOOH	+ O
Δ m/z (± 0.01 amu)	- 18.01057	- 43.98983	- 46.00548	+ 15.99492
Matching pairs	12	1	11	1

Ex. 1. Possible loss of H_2O



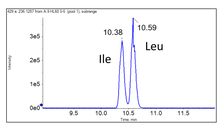
Ex. 2. Possible loss of HCO_2H



● Isobaric species

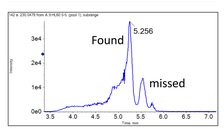
LC separation is crucial for isobaric metabolites, such as Ile and Leu.

The fact that features can be missed requires careful data review



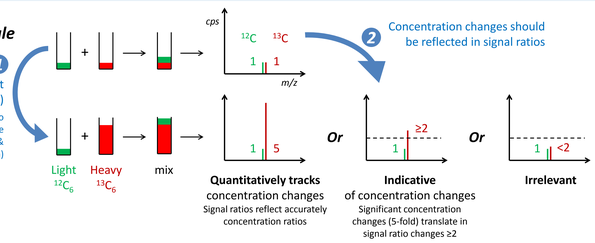
● "Missed" features

Close, unresolved peaks were found to be a cause of missed peak pairs



C Quantitative evaluation of results

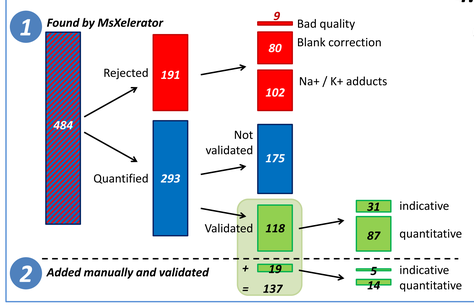
1. Rationale



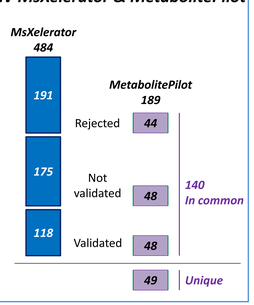
2. Validation criteria of relevant features

- All samples (7 different amount ratios from 1:1 to 5:5, 5 different series) were analyzed by LC-MS and the peak areas of the 293 pair list were integrated (some other pairs were manually added to the list).
- ***Pairs indicative of concentration changes*** ("indicative"):
  - A 5-fold concentration change: - is considered as representative of a significant biological change. - should lead to ≥ 2-fold increase or decrease in signals.
  - Criteria:
    - Concentration increase: the ratios 5:1 and 1:5 should be ≥ 2 times higher than 1:1
    - Concentration decrease: the ratios 5:1 and 1:5 should be ≥ 2 times lower than 5:5
    - CV of peak area ratios ≤ 30% for those samples
- ***Pairs that quantitatively track concentration changes*** ("quantitative")
  - Also indicative of concentration changes (i.e. meet criteria of previous category)
  - Peak area ratios should equal the sample amount ratio, 30% (max. 1 outlier in the 5 series)
  - CV of peak area ratios ≤ 30% for all samples (max. 1 outlier in the 5 series)

3. Final results



4. MsXelerator & MetabolitePilot



Conclusions & perspectives

Conclusions :

- A semi-targeted metabolomics workflow has been developed for the differential analysis of amine-containing metabolites
- The strategy takes advantage of selective reagents, available in light ($^{12}C_6$) and heavy ($^{13}C_6$) forms
- The differential labeling strategy allows for the detection unknown metabolites. From HL60 cells, 484 possible metabolites have been reported by MsXelerator.
- Relative quantitation allowed to track changes of concentration of possible metabolites in biological extracts. 137 possible metabolites were found to exhibit significant signal variation when their concentrations change

Perspectives :

- Evaluation of peak height (instead of peak area) and blank correction within MsXelerator software
- Evaluation of automatically-triggered CID experiments possible on QqTOF-MS instrument for structure elucidation of metabolites with accurate mass MS/MS spectra
- Identify unknown metabolites (using synthetic standards, when possible)
- Labeling of carboxylic acids with aniline as labeling reagent (available similarly in a heavy and a light form)