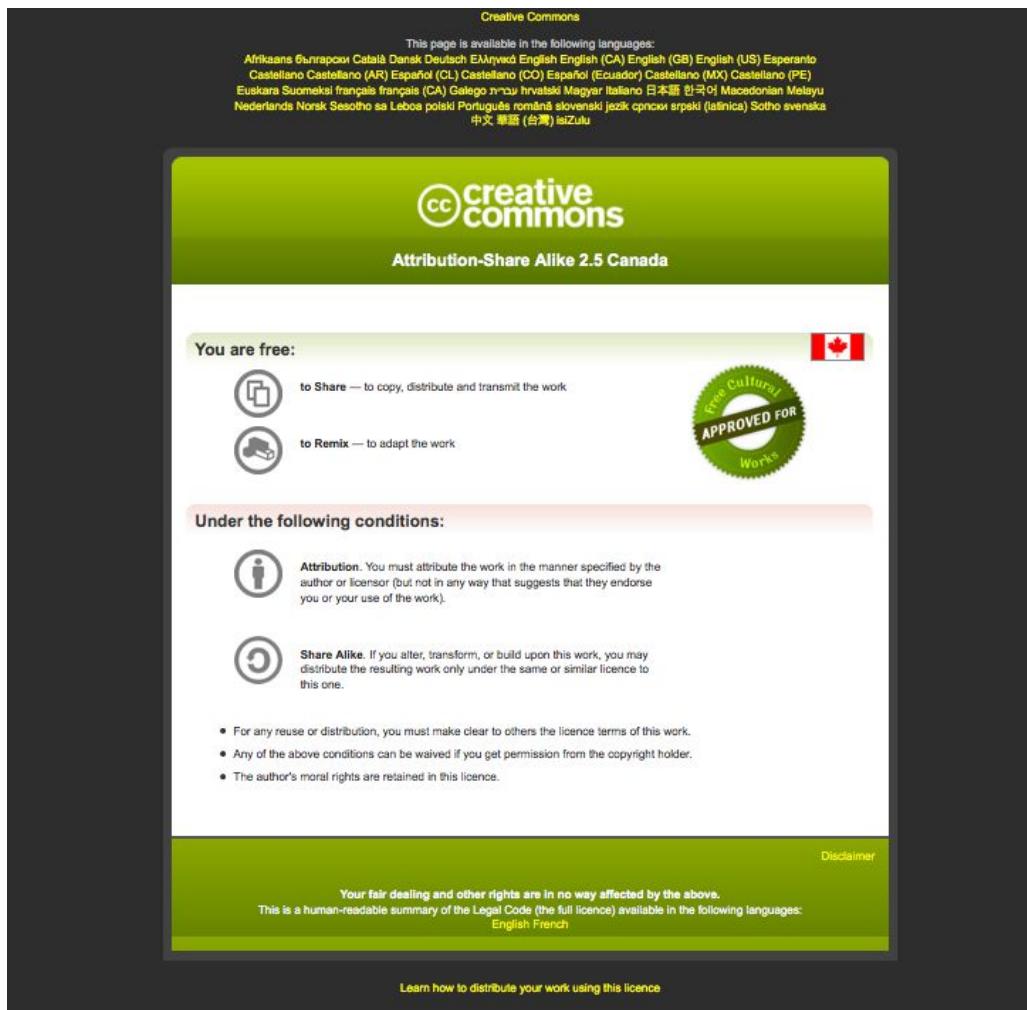




Canadian Bioinformatics Workshops

www.bioinformatics.ca

bioinformaticsdotca.github.io



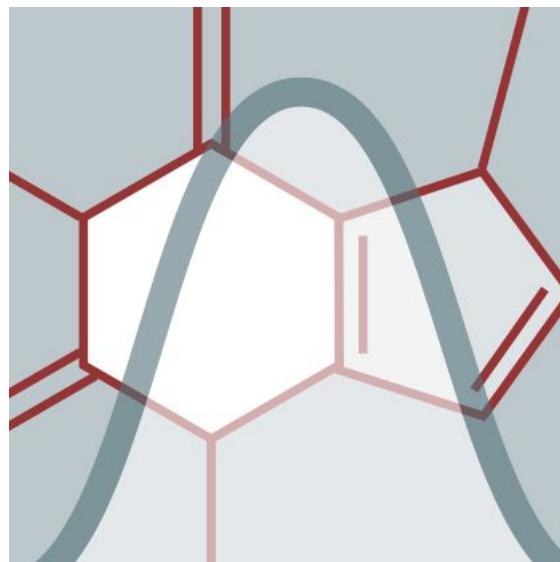
Targeted, Quantitative Metabolomics



David Wishart

Informatics and Statistics for Metabolomics

July 6-7, 2023



Schedule For July 6, 2023

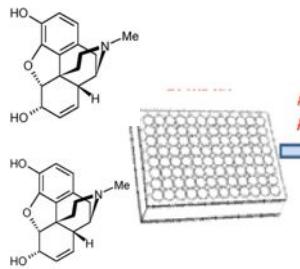
Time	Module
8:00 (MST)/10:00 (EST)	Arrival & Check-in
8:30 (MST)/10:30 (EST)	Welcome (Nia Hughes)
9:00 (MST)/11:00 (EST)	Module 1: Introduction to Metabolomics (David Wishart)
10:30 (MST)/12:30 (EST)	Break/Lunch (45 min)
11:15 (MST)/13:15 (EST)	Module 2: Targeted, Quantitative Metabolomics (David Wishart)
12:15 (MST)/14:15 (EST)	Lunch/Break (45 min)
13:00 (MST)/15:00 (EST)	Module 3 (Lab): Quantitative Metabolomics (David Wishart)
15:00 (MST)/17:00 (EST)	Break (30 min)
15:30 (MST)/17:30 (EST)	Module 4: Databases for Biological Interpretation (David Wishart)
17:00 (MST)/19:00 (EST)	Finish

Learning Objectives

- Understand the differences between targeted and untargeted metabolomics
- Learn about the importance of quantification in metabolomics
- Learn how quantitative NMR metabolomics is done
- Learn how quantitative GC-MS metabolomics is done
- Learn how quantitative LC-MS metabolomics is done

Two Routes to Metabolomics

Targeted (Quantitative)



Add Internal Stds

Add Samples

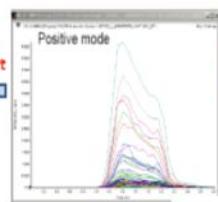
Protein Precipitation

Metabolite Extraction

MS Analysis

MS Analysis

QqQ or Qtrap MS
NMR
¹³C-MS



Asymmetric Dimethylarginine	0.36
L-arginine	2.14
Delta-1-hydroxybutyrate	50
Beta-alanine	125
All in all	15
Carnosine	0.85
Choline	1.2
Creatinine	0.5
Folate	0.025
Ketone	3450
Glucose	3450



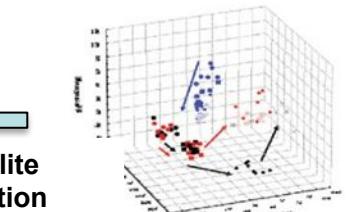
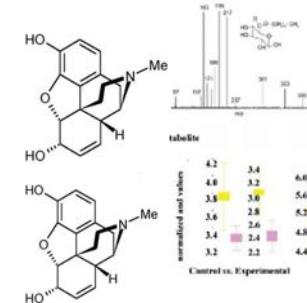
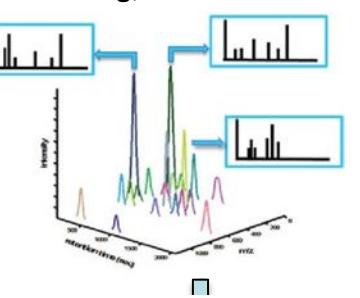
Untargeted (Non-Quantitative)



Liquid Chromatography
High Res Mass Spectrometry

Clustering, Peak detection

Data acquisition



Targeted vs. Untargeted

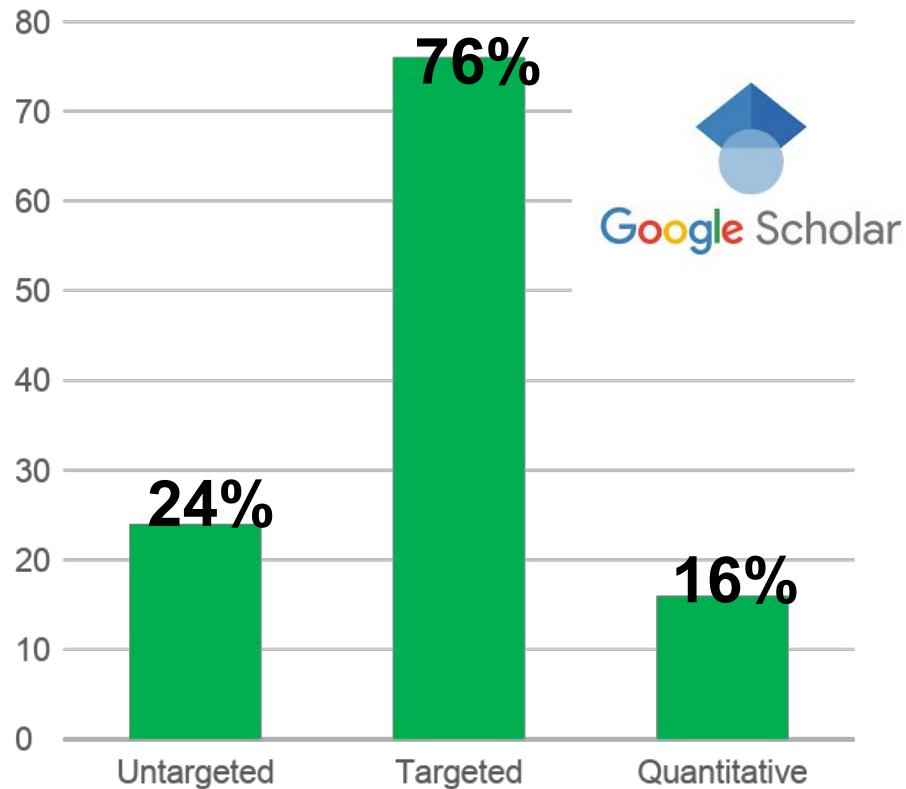
- Defined coverage of a set of pre-selected molecules (10-1400)
- General to all platforms
- Ideal for hypothesis testing but also for discovery via good use of quantification
- **Focus on absolute quantification**
- Fast, good for automation and kits
- Highly standardized or standardizable
- Open-ended, broad metabolite coverage (>10,000 features)
- Specific to HRMS
- Ideal for compound discovery and exploration (hypothesis generation)
- **Limited to relative quantification**
- Not very fast or very automated
- Needs much more standardization

Myths About Targeted or Quantitative Metabolomics

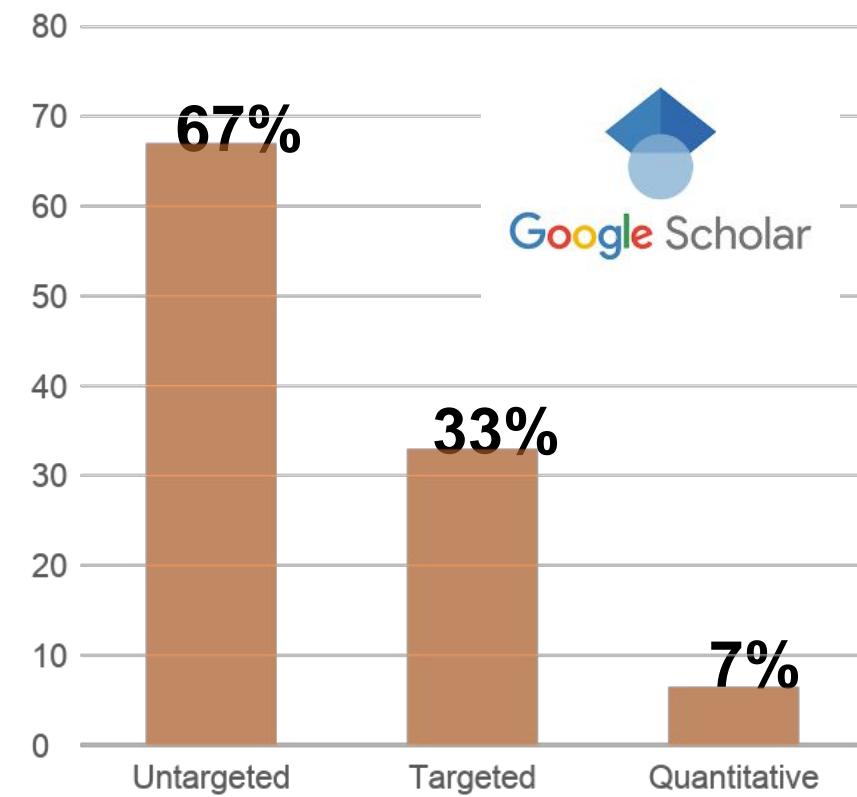
- ***Quantitative metabolomics is often viewed as:***
- More expensive
- More time-consuming
- Less sensitive
- Less comprehensive
- Less likely to lead to significant discoveries than untargeted metabolomics
- **As a result:** quantitative metabolomics is often relegated to the minor role of “confirmation” while untargeted metabolomics gets the starring role of “discovery”

Trends in Metabolomics

Metabolomics Papers
2012



Metabolomics Papers
2022



Why The Shift To Untargeted?

- Appeal of "discovering" a new compound or biomarker
 - Average of 1-2 new compounds ID'd per year or 1-2 for every 10,000 published metabolomics studies
- Belief that more compounds can be identified by untargeted than by targeted
 - Targeted now averages 500+ compounds ID'd while untargeted averages <100 compounds per study
- Faster and easier than targeted methods
 - Untargeted methods are 5-50X slower

Why The Shift To Untargeted?

- Untargeted methods are cheaper to run
 - Targeted assays can be run for <\$20/sample, untargeted assays usually cost >\$200/sample due to high data processing and informatics costs
- More opportunity for software development and software innovation
 - Nothing is standardized, moving target for software
- More likely to find novel, patentable biomarkers via untargeted methods
 - Can't patent natural chemicals, most biomarkers are found and assessed by concentration differences, FDA requires precise, quantitative measurements

What's Possible With Targeted Metabolomics

- NMR-based metabolomics (~50-200 metabolites identified and quantified with μM sensitivity)
- GC-MS based metabolomics (~20-120 metabolites identified and quantified with $<\mu\text{M}$ sensitivity)
- DI-MS based metabolomics (150-400 metabolites identified and quantified with nM sensitivity)
- LC-MS based metabolomics (300-800 metabolites identified and quantified with nM sensitivity)
- Lipidomics (3000 lipids identified and semi-quantified with nM sensitivity)

Definition of Analytical Chemistry

- A branch of chemistry that deals with the **quantitative determination of the chemical components of substances and mixtures**

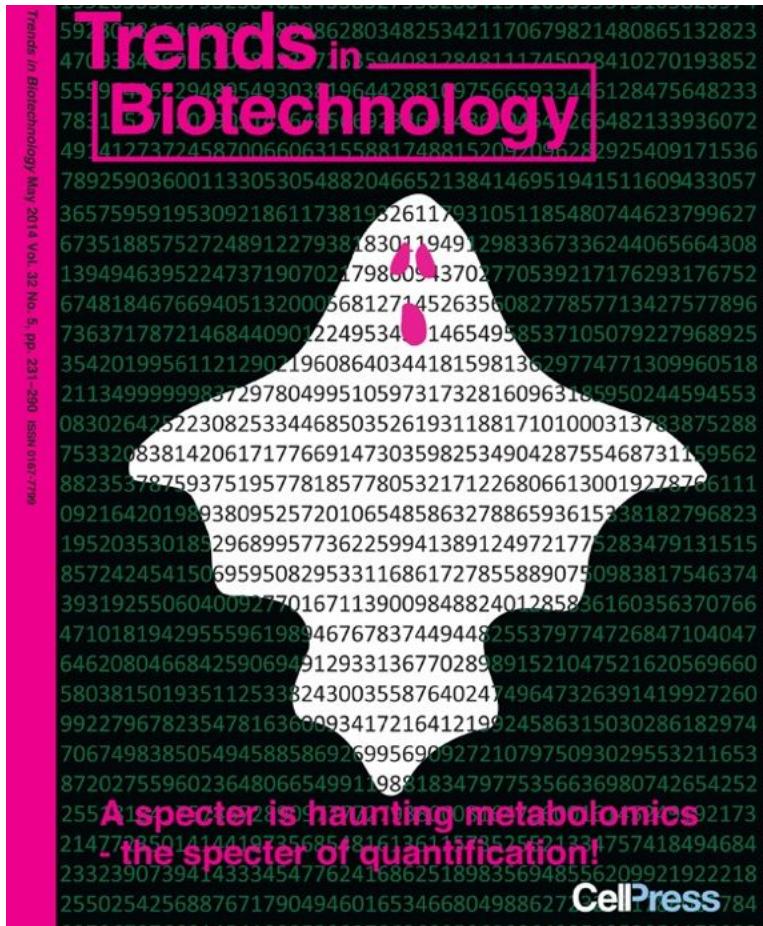


Why The Shift Away From Quantification?

- We have forgotten the definition of analytical chemistry
 - Quantify, quantify, quantify
- We have been influenced by “proteomics influencers” and proteomics refugees
 - MS-based proteomics workflows and concepts have been widely adopted by MS-based metabolomics labs
- Belief that medically useful biomarkers are qualitative measures not quantifiable entities
 - Just plain wrong

Quantification & Metabolomics

- This problem of quantification in metabolomics was identified as early as May 2014 by this article by S. Noack and W. Wiechert
- The field **MUST** become more quantitative if findings are to be translated to practical applications in human health and biomedicine



Noak & Wiechert, TIB May 2014

Advantages of Quantification

- It's reproducible
- It's verifiable
- It's universal
- The units don't change over time
- It allows direct comparisons to standardized (healthy) reference values
- *Avoids the need to always have 100s of healthy “control” samples in every study*



Why Quantification is Important for Metabolomics

- All the major (most cited) metabolomics discoveries for the past 15 years have involved measuring concentration changes of well-known metabolites
 - TMAO and cardiovascular disease (PMID: 21475195; PMID: 23614584)
 - BCAAs and AAAs in diabetes (PMID: 19356713; PMID: 21423183)
 - Host-gut microbiota/metabolism interactions (uremic toxins) (PMID: 19667173, PMID: 22674330; PMID: 31434538)
 - Oncometabolites (2-hydroxyglutarate, fumarate, succinate, lactate) (PMID: 19935646; PMID: 23747014; PMID: 23999438)
 - Immunometabolism (succinate, glucose, leucine, arginine) (PMID: 27396447; PMID: 23535595)
 - Biomarkers of food intake (proline betaine, 3-MeHis, TMAO) (PMID: 20573794; PMID: 28122782; PMID: 24760973)
 - Microbial metabolites and neurological disease (cresol-sulfate, indoxyl sulfate, HPPA) (PMID: 20423563; PMID: 30823925)

Quantitative Metabolomics Made Simple

- **Option 1 – do it yourself (requires LC-MS, NMR or GC-MS instrument, isotopic standards or isotopic labeling agents [LC-MS], unlabeled standards [NMR or GC-MS], reference calibration curves, follow protocols in literature, in-house software)**
- **Option 2 – do it yourself with ready-to-use instruments (buy NMR FoodScanner, NMR B.I. LISA or B.I. QUANT, buy Sciex Lipidlyzer, buy Seahorse XF, buy a commercial clinical analyzer)**

Quantitative Metabolomics Made Simple

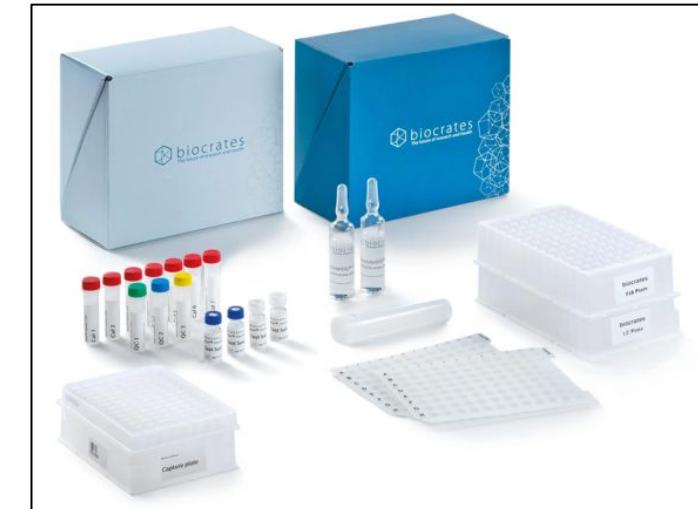
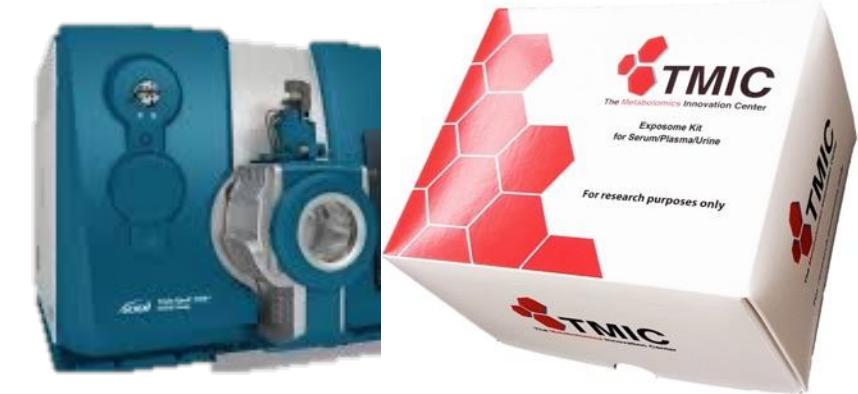
- **Option 3 – Send to academic labs offering fully quantitative metabolomics services**
 - <https://biocrates.com/certified-laboratories/>
 - TMIC (David Wishart, Phil Britz-McKibbin, Dajana Vuckovic, Jeff Xia)
 - Broad Institute (Clary Clish)
 - University of Washington (Dan Raftery)
 - Beaumont Health, Michigan (Stewart Graham)
 - UNC Chapel Hill (Susan Sumner)
 - Duke NCDRC (Chris Newgard)
 - WCMC-Targeted Assays (Oliver Fiehn)

Quantitative Metabolomics Made Simple

- **Option 4 – Send to commercial labs (Metabolon [19 quantitative assays], Biocrates, Chenomx, Nightingale Research Services, Metware Biotech)**
- **Option 5 – Buy and run metabolomics kits and/or methods from vendors (Biocrates [LC-MS], MetabolomiX [LC-MS, GC-MS, NMR], iMethods [Sciex], Lipidomics + Splash-Mix [Sciex + P. Baker], Chenomx [NMR], various MS vendor white papers)**

LC-MS Kits

- Quantitative, kit-based, targeted MS metabolomics systems with hundreds of chemical standards, bundled software and ISO-level protocols
- Permit >640 compounds (soon 1200+) to be accurately measured
- Easy to use, training videos and ported to several labs already
- \$45-\$90/sample, 96 well format



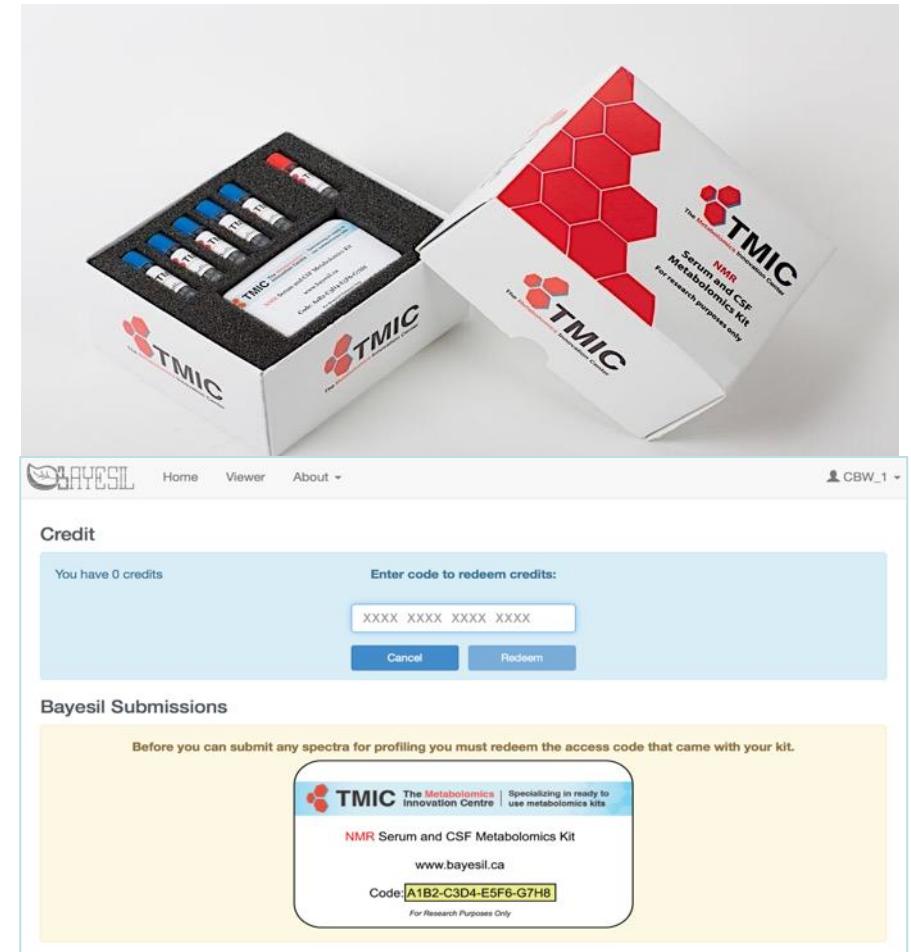
GC-MS Kits

- Quantitative, kit-based, targeted GC-MS metabolomics systems with >100 chemical standards, derivatization agents, bundled software and ISO-level protocols
- Accepts NetCDF or mzXML files
- 60 sec per spectrum
- 40-115 cmpds ID'd and quantified, 96% accuracy
- Optimized for blood, urine, saliva and CSF
- Requires careful sample preparation & derivatization



NMR Kits

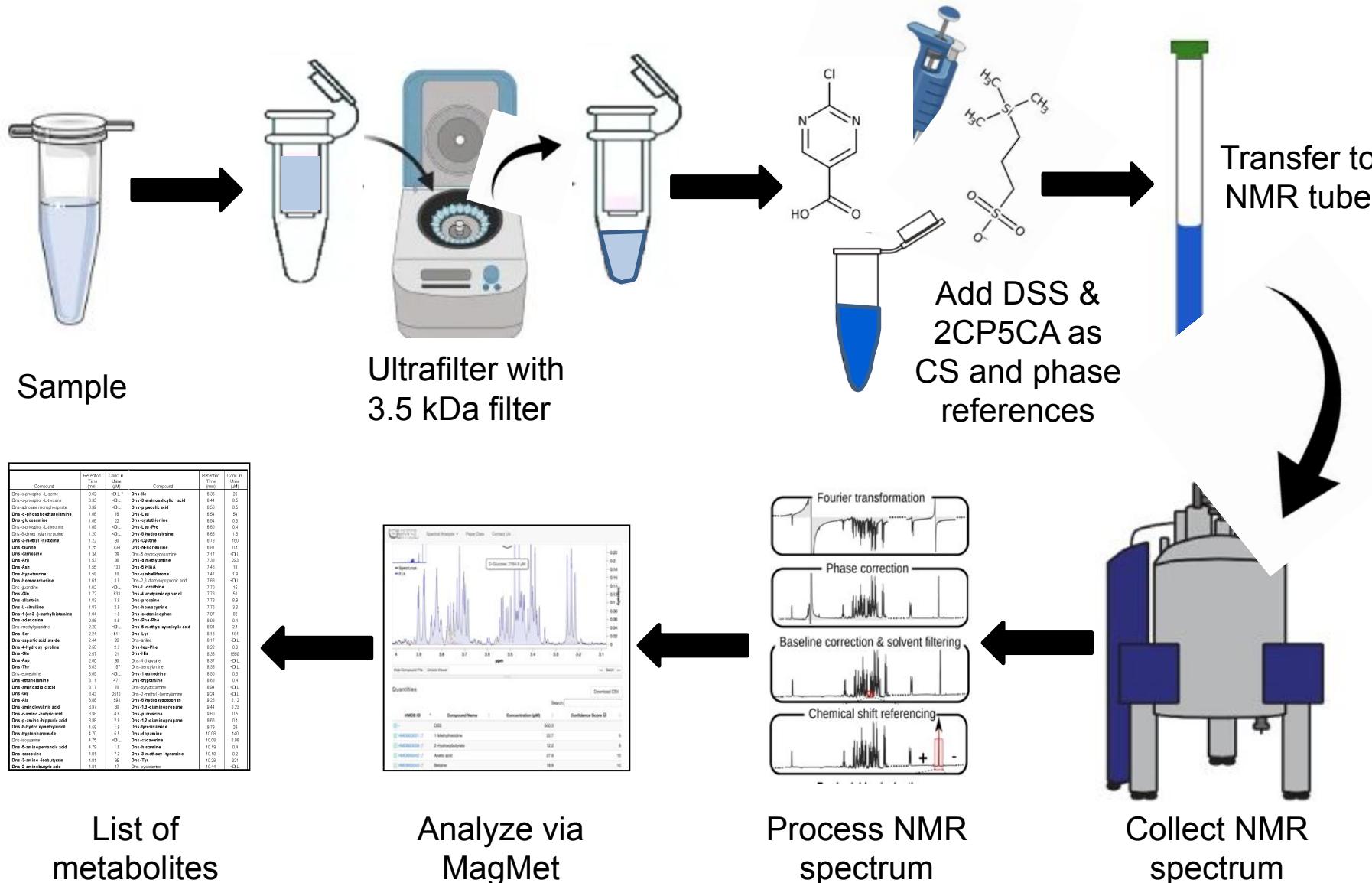
- Developed to help standardize NMR metabolomics in TMIC and to help other labs standardize NMR-based metabolomics
- Significantly improved performance (batch analysis, faster, improved accuracy)
- Interactive spectral editing
- Includes all required reagents, standards and software access code for processing 100 samples & detecting 60+ cmpds in serum, plasma or fecal water
- Price: \$10/sample, 50 samples/kit



rmandal@ualberta.ca

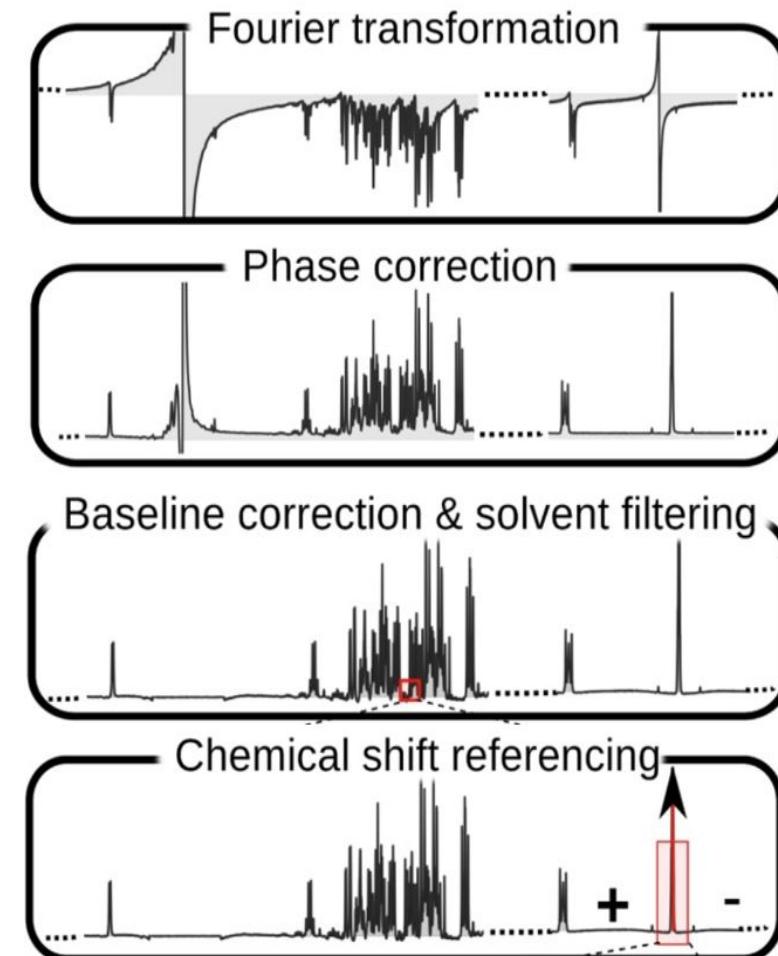
NMR Metabolomics

NMR Kit - Assay Overview



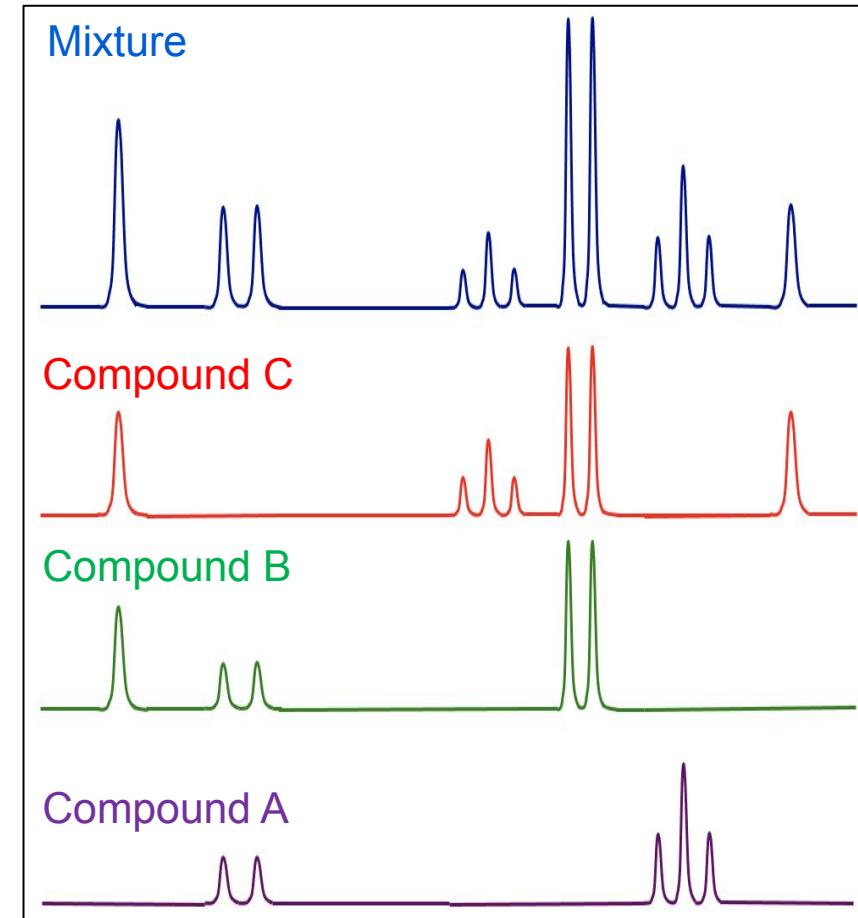
NMR Spectral Processing

- NMR spectra usually have to be manually adjusted where the FID is Fourier transformed, the spectrum is phased (peaks point upwards and have a symmetric shape), then the baseline is flattened and the large water peak is removed
- Additionally, the chemical shift reference (0 ppm) has to be defined and set



NMR Spectral Deconvolution

- Spectral deconvolution is a method for identifying and quantifying individual compounds in a mixture by decomposing mixture spectra into individual pure compound spectra
- Matches peaks to a known set of peaks (from pure compounds) in a pre-compiled spectral database



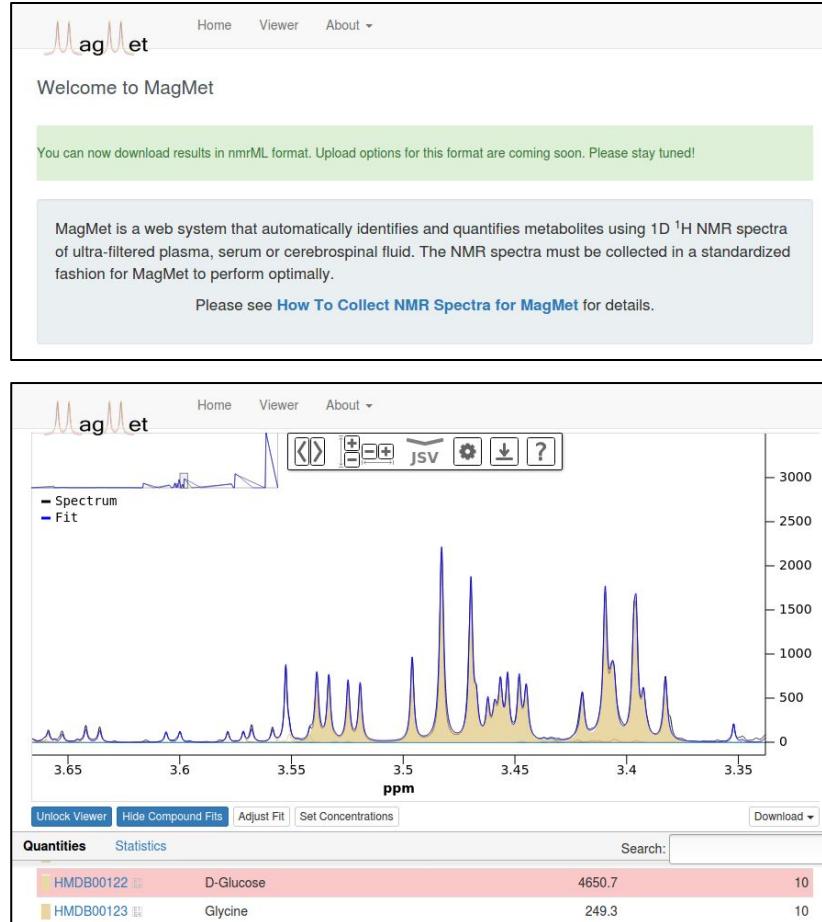
NMR Spectral Deconvolution

- Can be done manually, semi-automatically or fully automatically
- First methods were done manually where individuals had to process the spectra by hand, a large spectral library was provided and users then had “guess and check” which pure compound spectra matched by clicking, dragging and iteratively adjusting the spectra to match the peaks in the mixture being studied

Advantages of Automation

- **10-20 times faster than manual methods (much higher throughput)**
- **Precision and recall is often > 95%**
- **Permits unsupervised analysis and overnight data processing (prevents burn-out)**
- **Consistent and reproducible (not prone to user bias or user errors)**
- **Detects and quantifies signals that humans cannot easily detect or quantify**

MagMet - For NMR Spectral Deconvolution



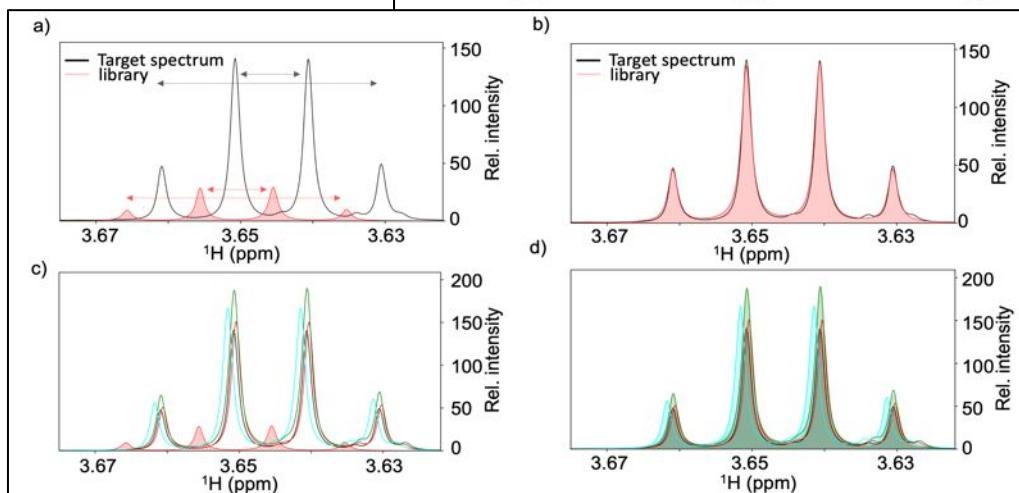
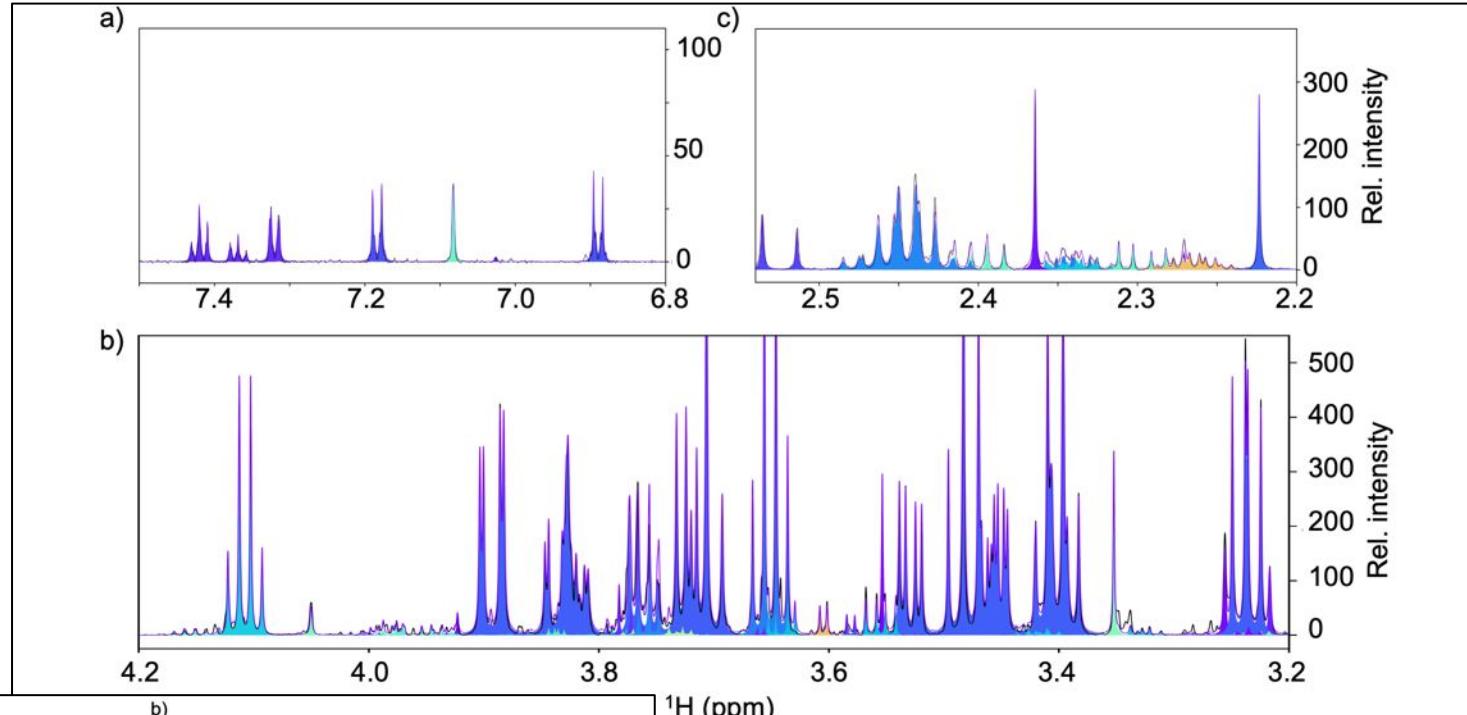
- MagMet – Magnetic Resonance for Metabolomics – Replaces Bayesil
- Offers more flexibility than Bayesil (which was limited to analyzing blood and CSF)
- Uses a combination of machine learning and expert-based rules to perform fitting and pattern finding
- Fully automated phasing, referencing, water removal, baseline correction, peak convolution, ID and quantification
- Detects & quantifies 55-60 serum metabolites in ~10 minutes
- Has been adopted to wine, beer, juice, milk, fermentation products

<https://teaching.magmet.ca>

MagMet Spectral Fitting

Performs peak integration for concentration determination

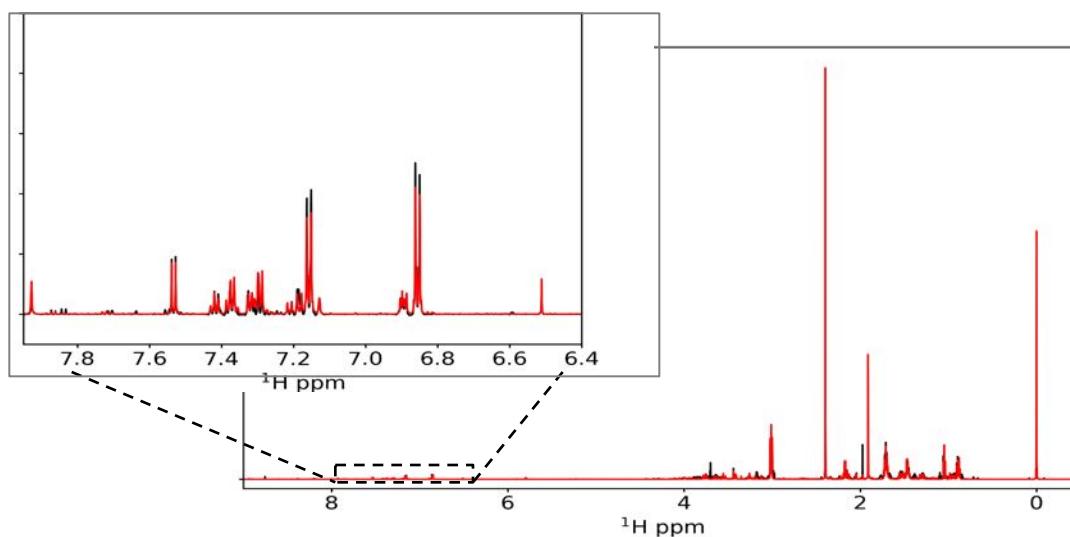
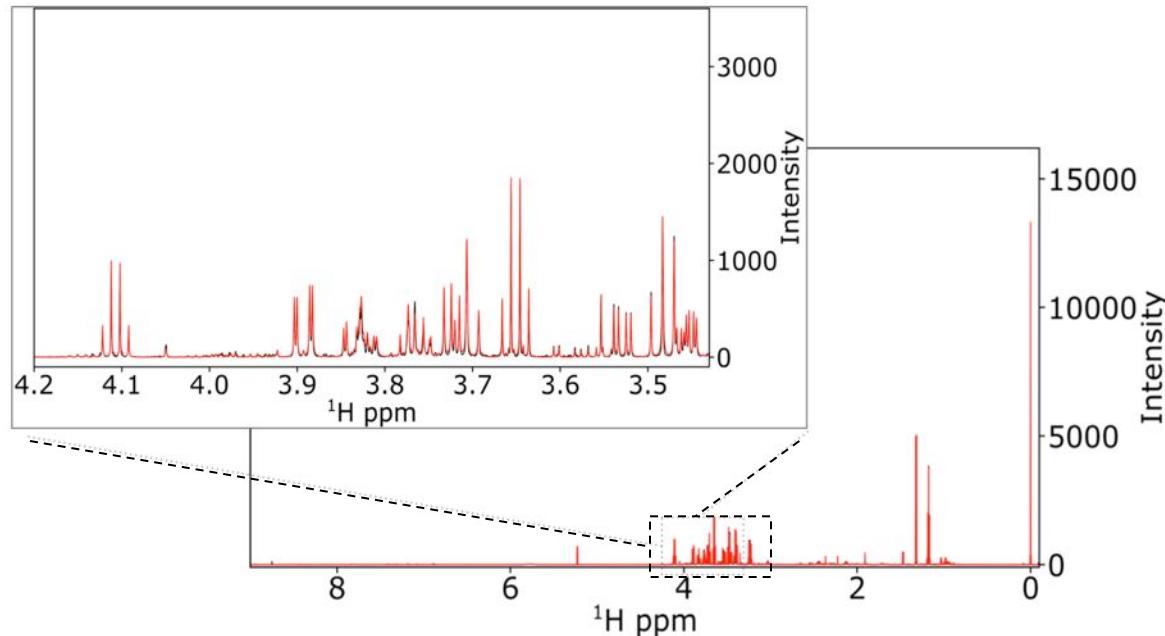
Colors indicate compound ID confidence



Iterative fitting performed with peak shift matching (first) and then peak intensity matching

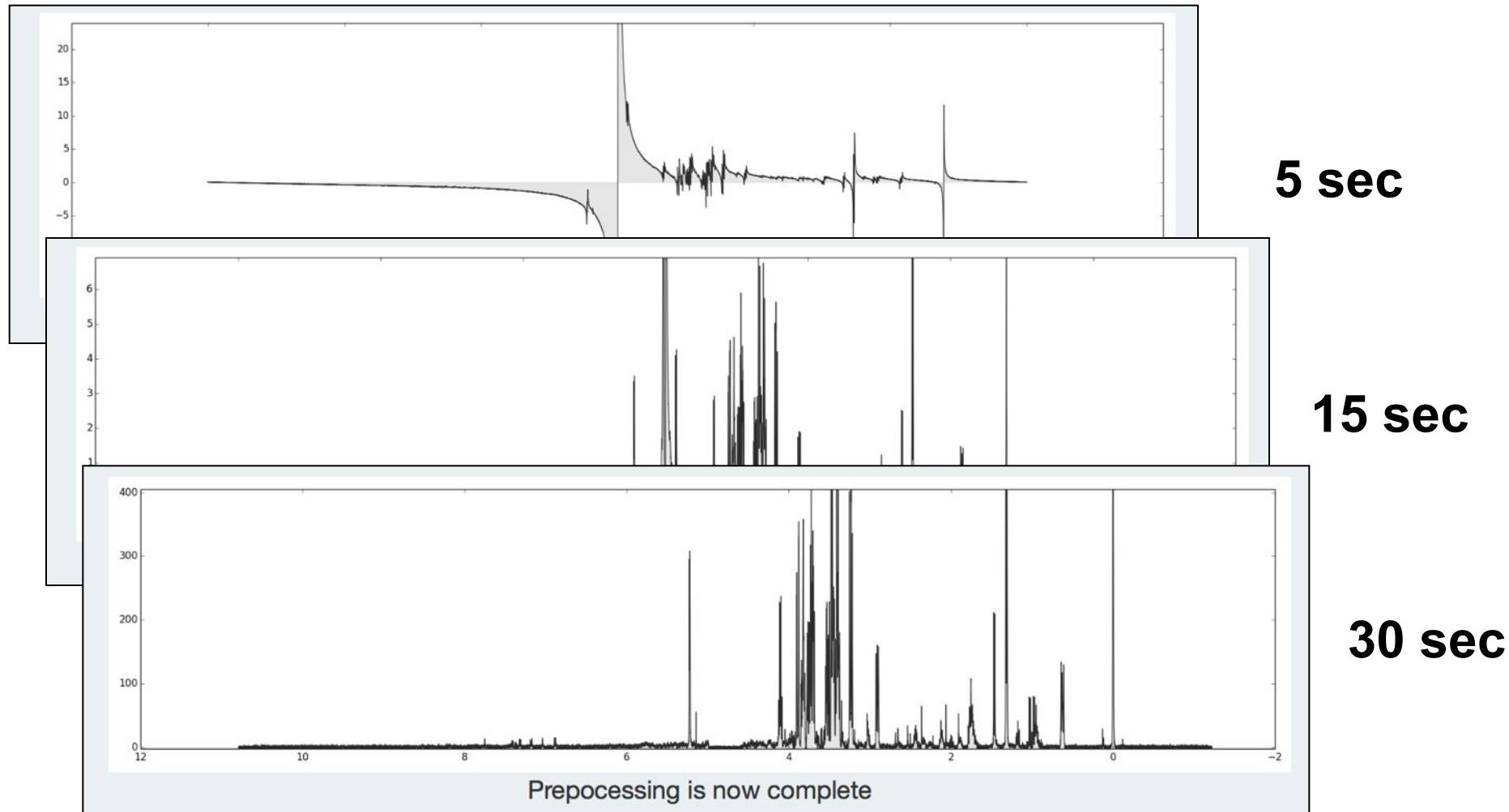
MagMet Performance

MagMet (serum)
57 metabolites ID'd
CV < 8%
4 min. 59 seconds

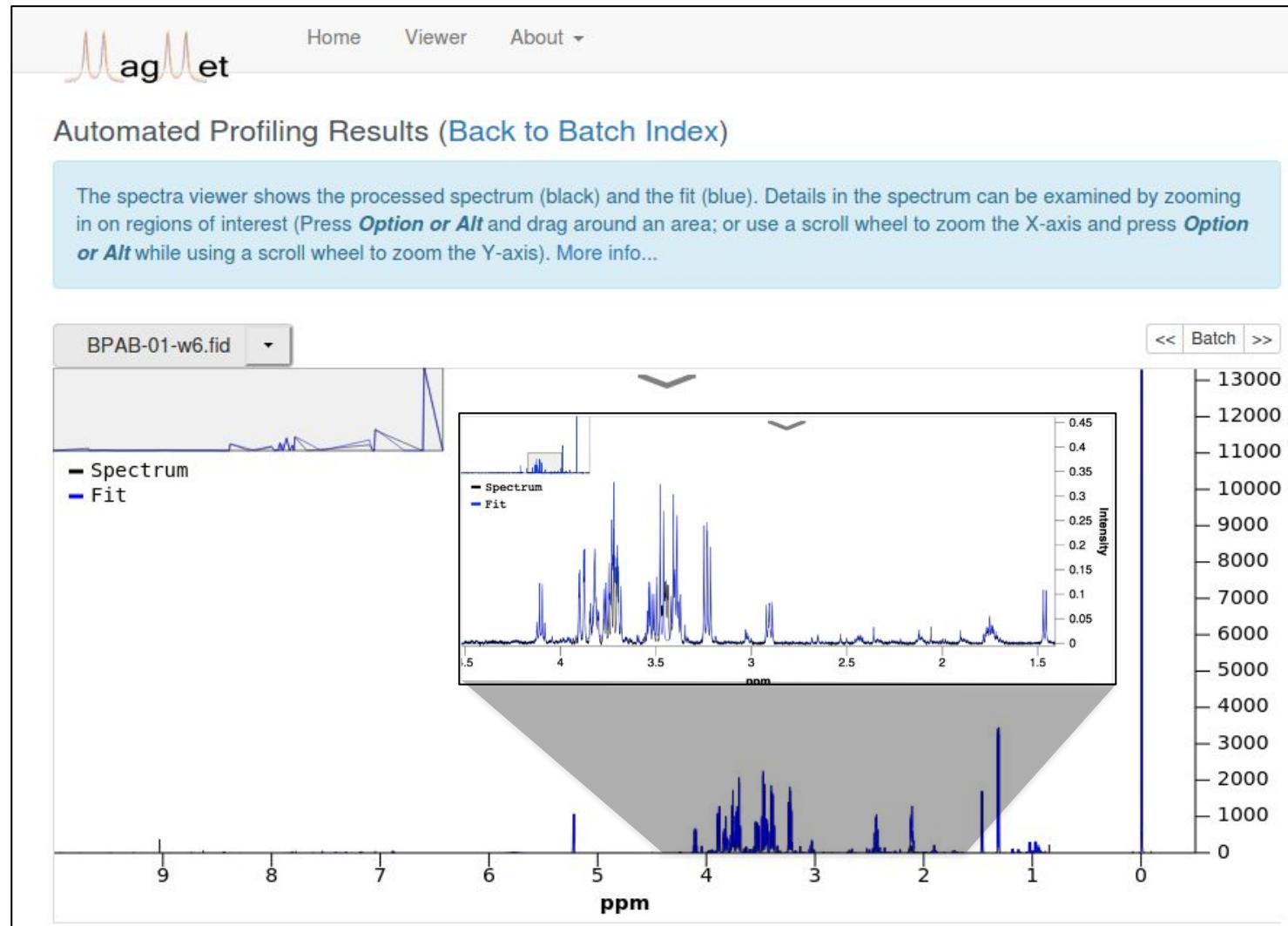


MagMet (fecal water)
72 metabolites ID'd
CV < 10%
6 min 13 seconds

MagMet in Operation



MagMet in Operation



Metabolite ID and Quantification by MagMet

Quantities	Settings	Statistics	Search:
HMDB ID	Compound Name	Concentration (μ M)	Confidence Score
-	DSS	1000.0	
HMDB00001	1-Methylhistidine	0.0	5
HMDB00008	2-Hydroxybutyrate	24.5	9
HMDB00042	Acetic acid	55.3	10
HMDB00043	Betaine	35.1	10
HMDB00060	Acetoacetate	0.0	8
HMDB00062	L-Carnitine	41.5	9
HMDB00064	Creatine	32.6	7
HMDB00094	Citric acid	123.9	10
HMDB00097	Choline	12.6	10
HMDB00108	Ethanol	78.0	9
HMDB00122	D-Glucose	6106.6	10

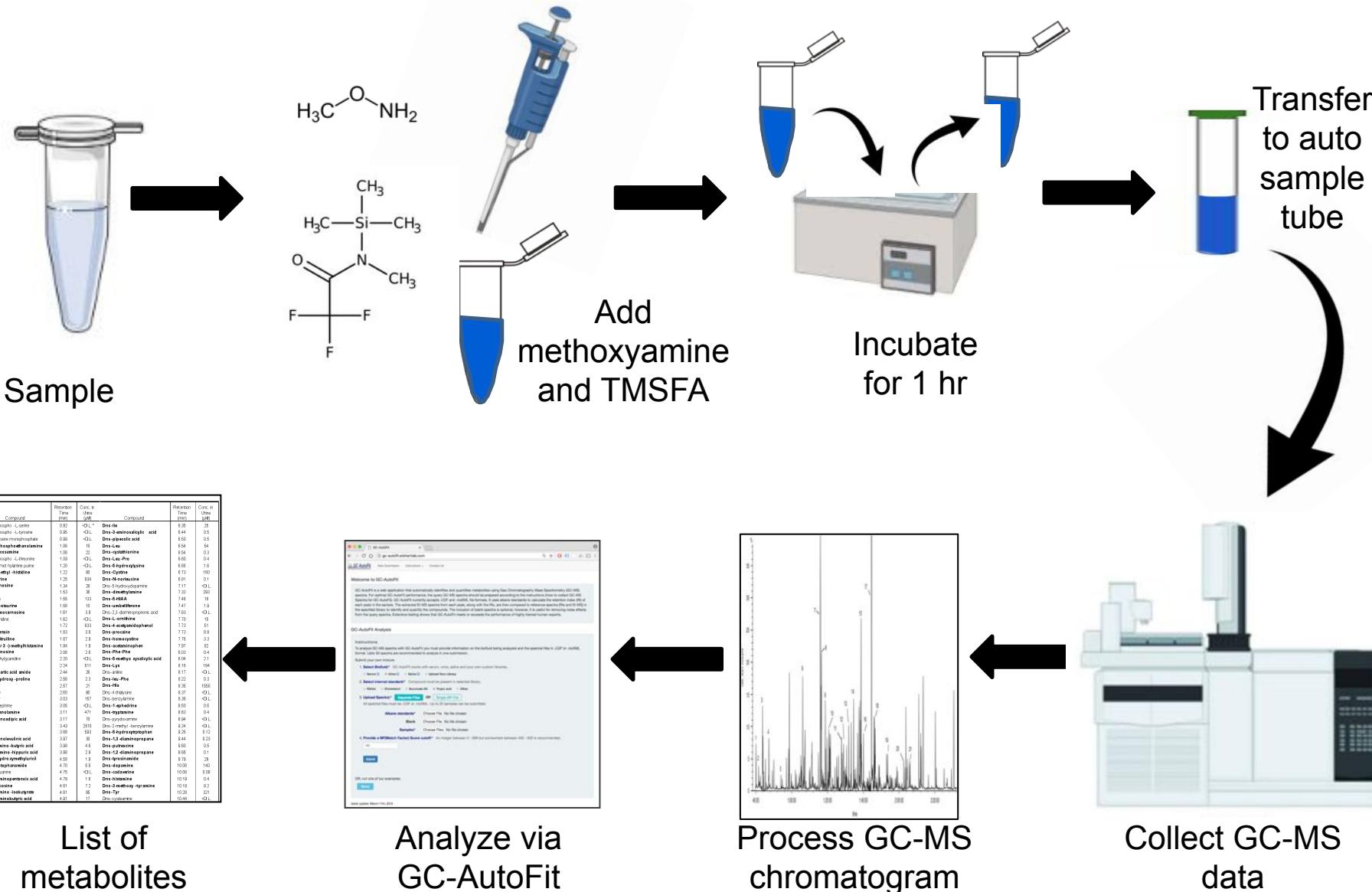
GC-MS Metabolomics

GC-MS Kits

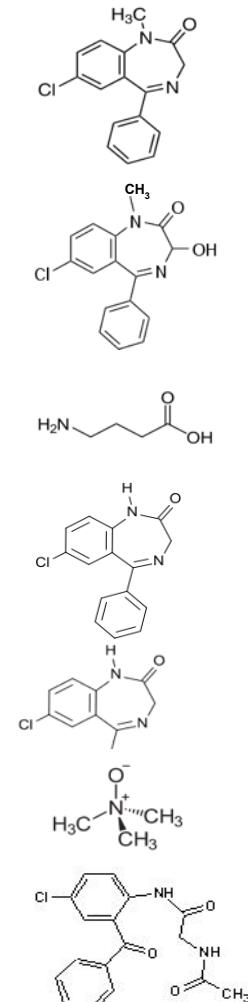
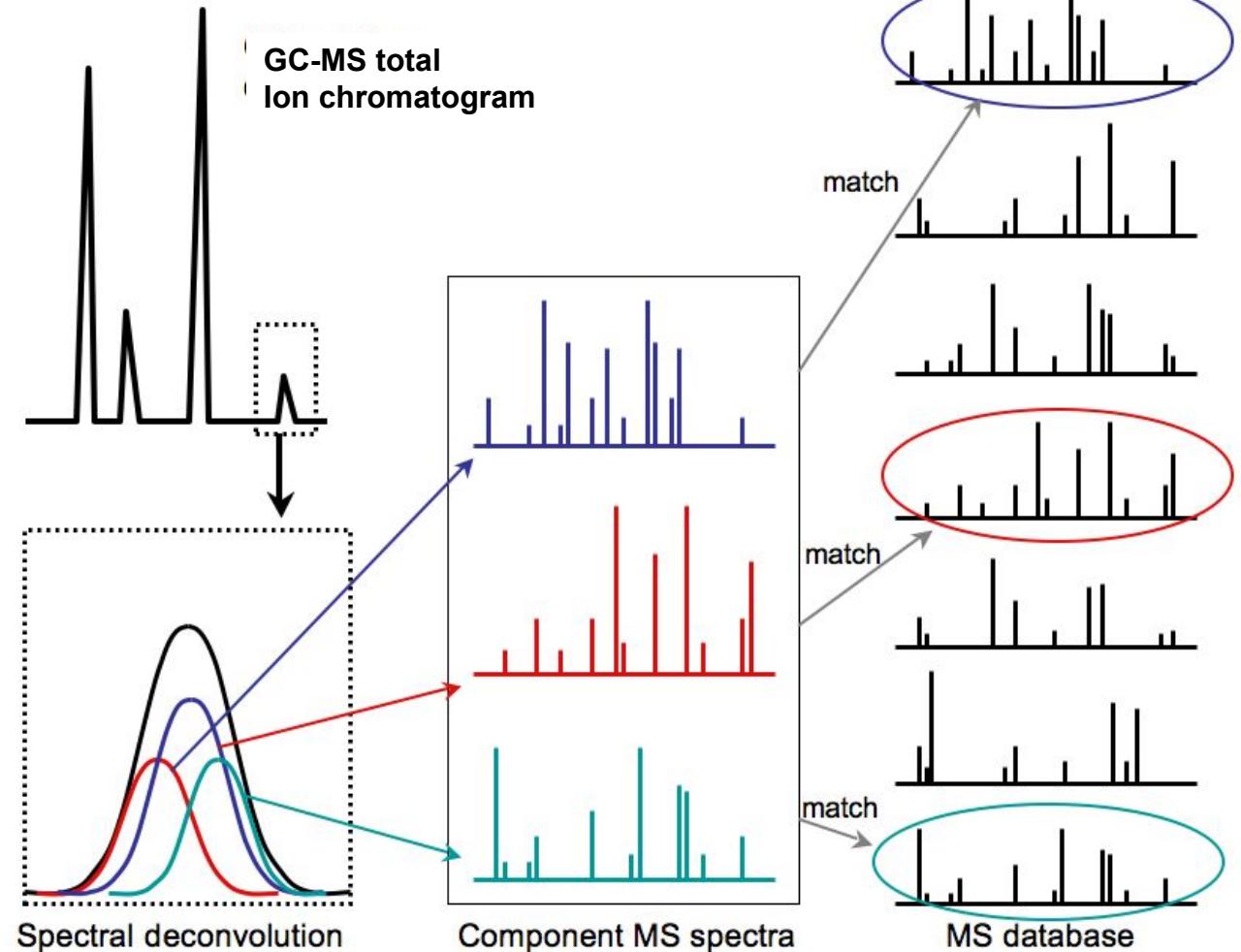
- Quantitative, kit-based, targeted GC-MS metabolomics systems with >100 chemical standards, derivatization agents, bundled software and ISO-level protocols
- Accepts NetCDF or mzXML files
- 60 sec per spectrum
- 40-115 cmpds ID'd and quantified, 96% accuracy
- Optimized for urine (other biofluids upcoming)
- Requires careful sample preparation & derivatization



GC-MS Kit - Assay Overview

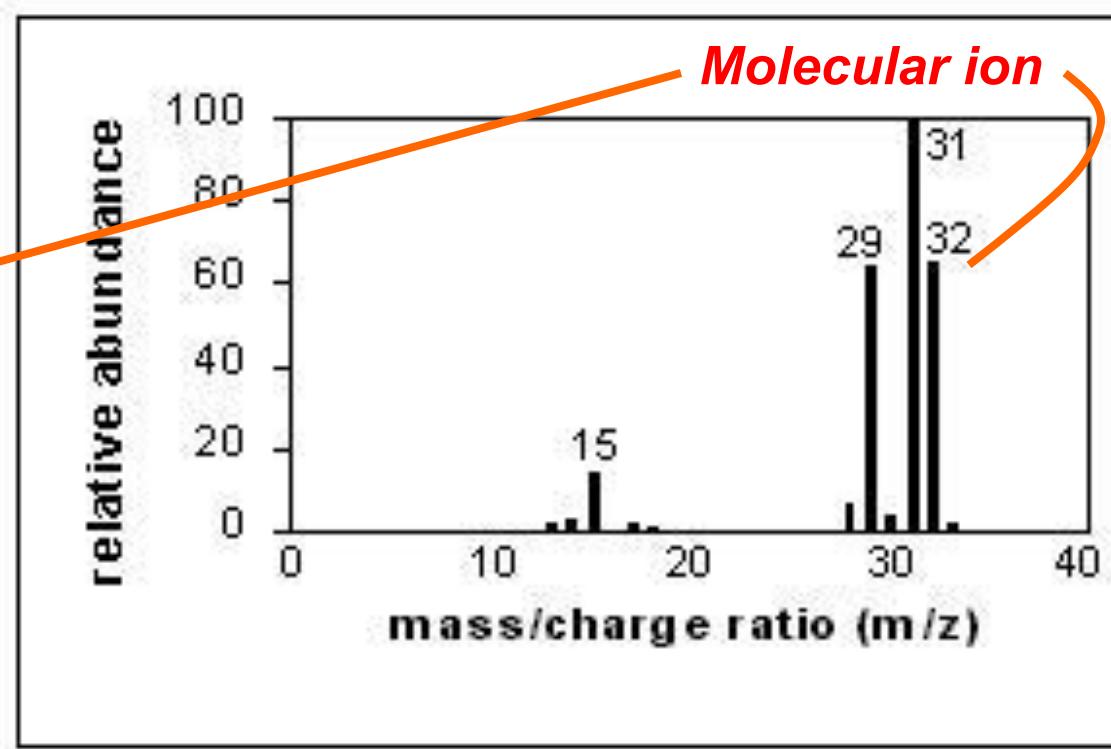


Metabolite ID by GC-MS



Recall EI MS Generates Multiple Peaks

ions	m/z
CH_3OH^+	32
$\text{H}_2\text{C}=\text{OH}^+$	31
$\text{HC}\equiv\text{O}^+$	29
H_3C^+	15



EI Breaks up Molecules in Moderately Predictable Ways

Recall GC-MS Analytes are Derivatized

Derivatization reaction	Number of groups	Mass increment
TMS	1	72
$\text{R-O-H} \longrightarrow \text{R-O-Si}-\begin{array}{c} \\ \text{I} \end{array}$	2	144
	3	216
	4	288
$\text{R-N}-\begin{array}{c} \\ \text{H} \end{array}-\text{H} \longrightarrow \text{R-N}-\text{Si}-\begin{array}{c} \\ \text{H} \end{array}-\begin{array}{c} \\ \text{I} \end{array}$	5	360
	6	432
TBDMS	1	114
$\text{R-O-H} \longrightarrow \text{R-O-Si}-\begin{array}{c} \\ \text{I} \end{array}-\begin{array}{c} \\ \text{I} \end{array}$	2	228
	3	342
	4	456
$\text{R-N}-\begin{array}{c} \\ \text{H} \end{array}-\text{H} \longrightarrow \text{R-N}-\text{Si}-\begin{array}{c} \\ \text{H} \end{array}-\begin{array}{c} \\ \text{I} \end{array}-\begin{array}{c} \\ \text{I} \end{array}$	5	570
	6	684
MO Methoxyamine	1	29
$\text{R}-\text{C}(=\text{O})-\text{R}_1 \longrightarrow \text{R}-\text{C}-\text{R}_1-\begin{array}{c} \text{O}- \\ \parallel \\ \text{N} \end{array}$	2	58
	3	87
	4	116

Metabolite ID by GC-MS

- GC-MS is often best for identification of amino acids, organic acids, sugars, fatty acids and molecules with MW<600
- GC has higher resolution and better reproducibility than LC
- EI-MS is more standardized than soft ionization methods, so EI spectra are more comparable
- Observed EI-MS spectra are compared to a database of experimental EI-MS spectra and scored by similarity

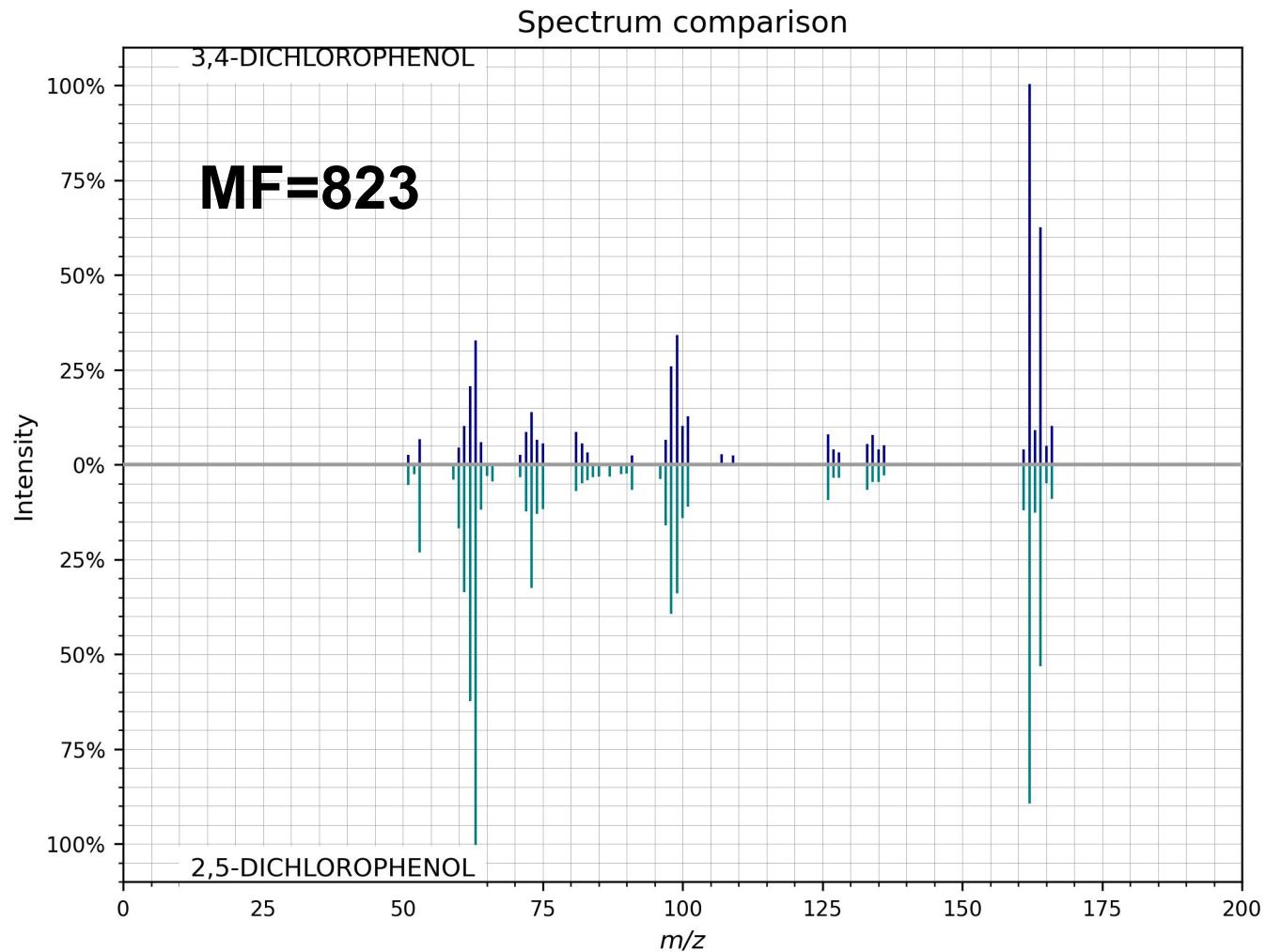
Match Factor (MF)

- Measures the similarity of the MS spectrum of the query to the MS spectrum in the reference database
- Defined as the normalized dot product of the query and the reference spectra

$$MF = \frac{1000 * (\sum_w M [I_{qry} - I_{ref}]^{1/2})^2}{\sum I_{qry} M * \sum I_{ref} M}$$

I_{ref} corresponds to the intensities of the reference spectra,
 I_{qry} corresponds to the intensities of the query spectra,
 M corresponds to the masses (m/z)
 w is a weighting term to penalize uncertain peaks

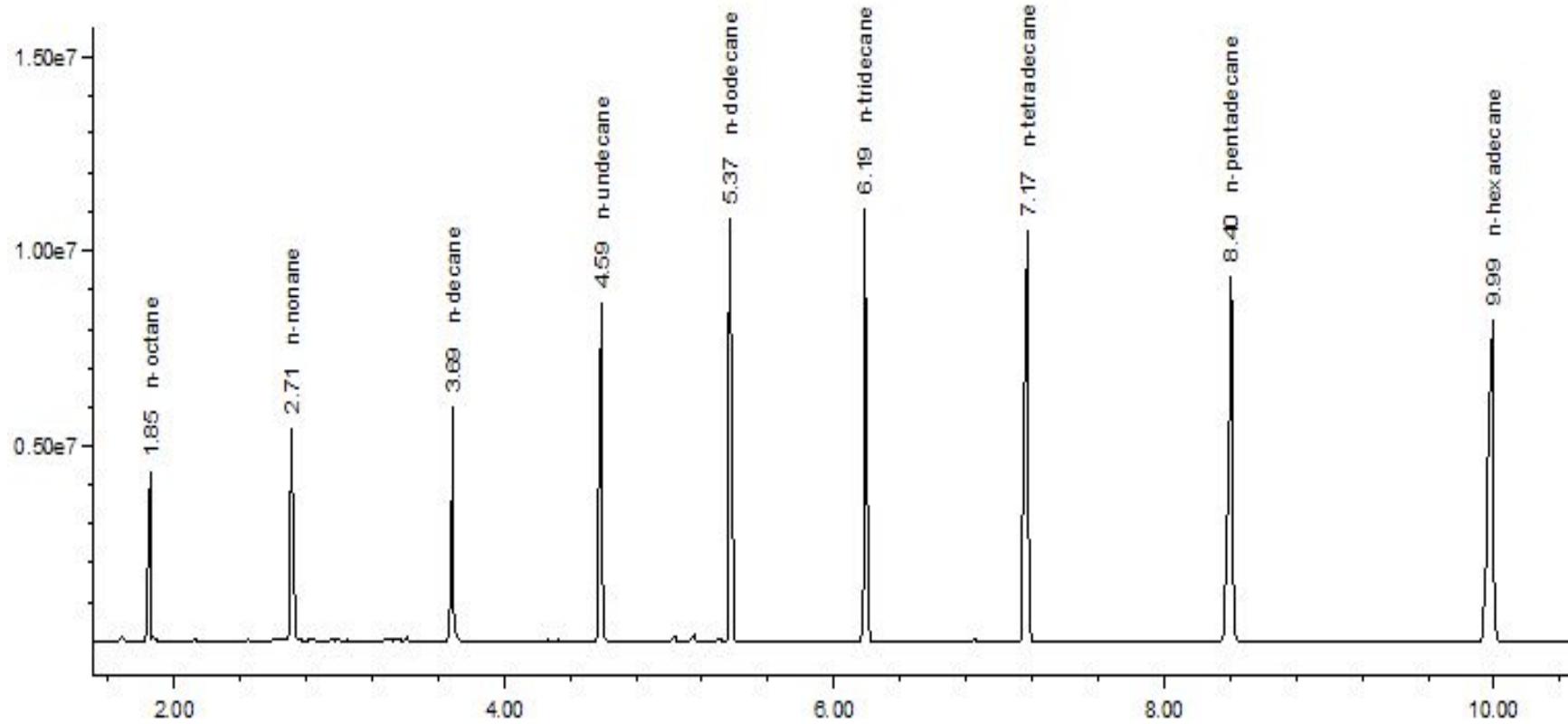
Match Factor Visualized



GC-MS Protocol

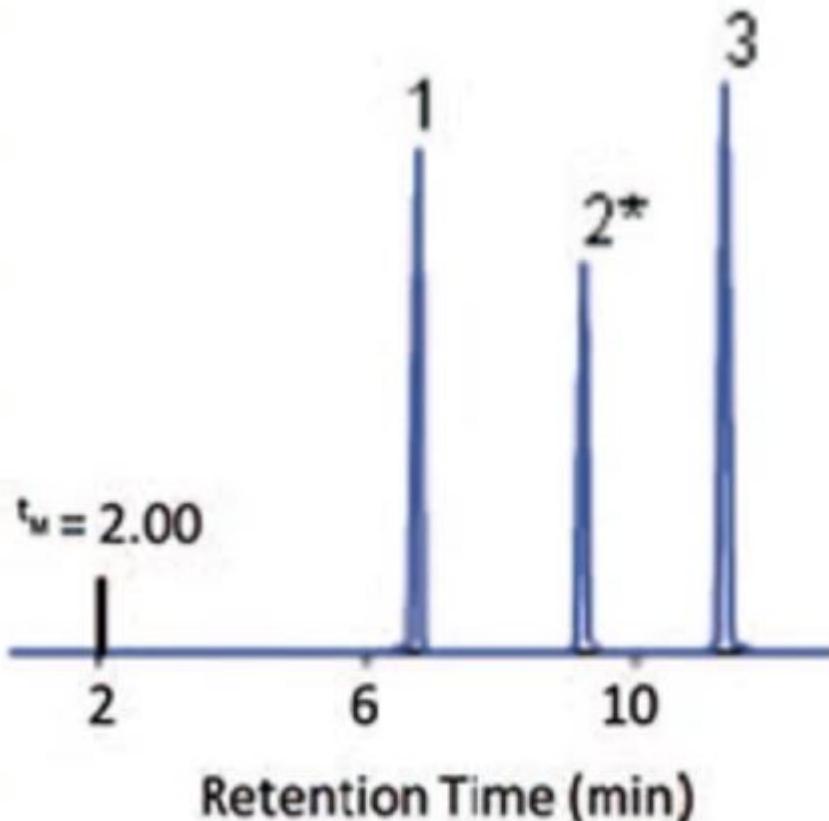
- Prepare a set of external n-alkane standards (8-9 n-alkanes spanning octane to hexadecane) and run as an external calibration standard
- Run a “blank sample” containing just the solvent and derivatization agents
- Run the sample of interest (under the same conditions as the blank)
- Create a calibration file using the n-alkane mixture (sets retention indices [RI's] to the standard values)

GC-MS Protocol



External n-alkane standard used for RI calculation

Retention Index



1. Hexane $t_R = 6.50$ min, $t'_R = 4.50$ min
2. Unknown $t_R = 8.25$ min, $t'_R = 6.25$ min
3. Heptane $t_R = 11.50$ min, $t'_R = 9.50$ min

$$I = 100(6) + 100 \left[\frac{\log 6.25 - \log 4.50}{\log 9.50 - \log 4.50} \right]$$

$$I = 644$$

$$I = 100 \left[n + (N - n) \frac{\log t'_r(\text{unknown}) - \log t'_r(n)}{\log t'_r(N) - \log t'_r(n)} \right]$$

where n is the number of carbon atoms in the *smaller* alkane

N is the number of carbon atoms in the *larger* alkane

$t'_r(n)$ is the adjusted retention time of the *smaller* alkane

$t'_r(N)$ is the adjusted retention time of the *larger* alkane

GC-MS Protocol

- Analyze the sample data file against the CAL(calibration)-file for the alkane mixture (sets and recalculates RI's using the n-alkanes)
- Get rid of any “false” positives by comparing the “blank” against the sample spectrum
- Search a spectral database for spectral matches using the match factor score and retention index similarity and display the results of the search

GC-AutoFit (Automated GC-MS)

The screenshot shows the 'GC-AutoFit Analysis' page. At the top, there's a 'Welcome to GC-AutoFit' message and a brief description of how it identifies metabolites using GC-MS spectra. Below this is a 'GC-AutoFit Analysis' section with 'Instructions' for uploading spectra in .CDF or .mzXML format. It includes fields for 'Select Biofluid' (radio buttons for Serum, Urine, Saliva, or Upload Your Library), 'Select internal standard' (radio buttons for Ribitol, Cholesterol, Succinate-D4, Tropic acid, or Other), and 'Upload Spectra' (checkboxes for Separate Files or Single ZIP File). A note says up to 30 samples can be submitted. There are also fields for 'Alkane standards' (Choose File) and 'Samples' (Choose Files). A 'Provide a MF(Match Factor) Score cutoff' input field is set to 400, with a note that an integer between 0-999 is recommended. A 'Submit' button is at the bottom. Below the main form, there's a section for examples with a 'Serum' button.

- Requires 3 spectra (sample, blank, alkane standards)
- Performs auto-alignment, peak ID, peak integration and concentration calculation
- Accepts NetCDF or mzXML files
- 60 sec per spectrum
- 40-116 cmpds ID'd and quantified, 96% accuracy
- Optimized for blood, urine, saliva and CSF
- Still requires careful sample preparation & derivatization

<http://gc-autofit.wishartlab.com>

Preparing GC-MS Spectral Files for GC-AutoFit

- Three types of input files
 - Alkane standard file (required)
 - e.g., Alkane.mzXML, ALKstd.mzXML
 - A Blank sample file (optional but recommended)
 - e.g., Blank.mzXML, Blk.mzXML
 - Sample files (required)
- File format conversion, if necessary
 - CDF and mzXML format are expected as an input file format
 - '.D' format, '.CDF' or '.mzXML' format
 - Conversion Software
 - ChemStation
 - ProteoWizard (<http://proteowizard.sourceforge.net>)

The screenshot shows the 'GC-AutoFit' web application. At the top, there are navigation links for 'New Submission', 'Instructions', and 'Library'. The main content area has a header 'Welcome to GC-AutoFit' with a brief description of the application's purpose: to automatically identify and quantify metabolites using Gas Chromatography Mass Spectrometry (GC-MS) spectra. Below this is a section titled 'GC-AutoFit Analysis' with instructions for submission. It states that to analyze GC-MS spectra, biofluid information and spectral files in .CDF or .mzXML format are required. Up to 30 spectra can be submitted. A note says that blank spectra are optional but useful for removing noise effects. Step 1, 'Select Biofluid:', has a radio button for 'Urine' which is selected. Step 2, 'Select internal standard:', has a radio button for 'Tropic acid'. Step 3, 'Upload Spectra:', offers options to upload separate files or a single ZIP file. Step 4, 'Provide a MF(Match Factor) Score cutoff:', has a text input field set to '400'. At the bottom are 'Submit' and 'Clear' buttons.

Uploading GC-MS Spectral Files

- Option: Individual files (A) or a zip file (B) or (C) an example file

(A)

3. Upload Spectra:
 OR

All spectral files must be .CDF or .mzXML. Up to 30 samples can be submitted.

Alkane standards* No file selected.

Blank No file selected.

Samples* No files selected.

(B)

3. Upload Spectra:
 OR

The ZIP file must contain an alkane standards spectrum (e.g. Alkane.mzXML, ALKstd.mzXML), an optional blank spectrum (e.g. Blank.mzXML, Blk.mzXML) and at least one sample spectrum. All the spectral files must be .CDF or .mzXML. Up to 30 samples can be submitted.

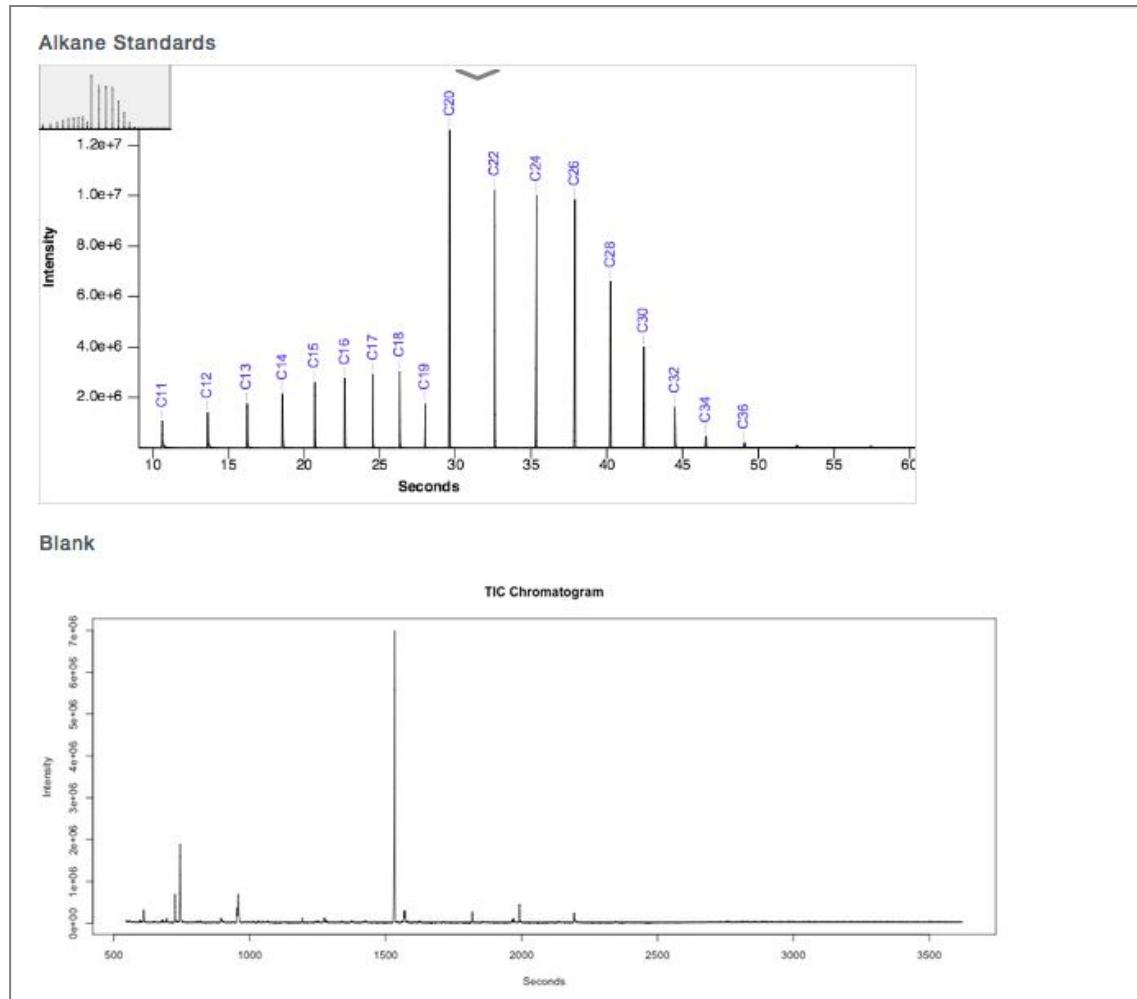
Zip file* No file selected.

(C)

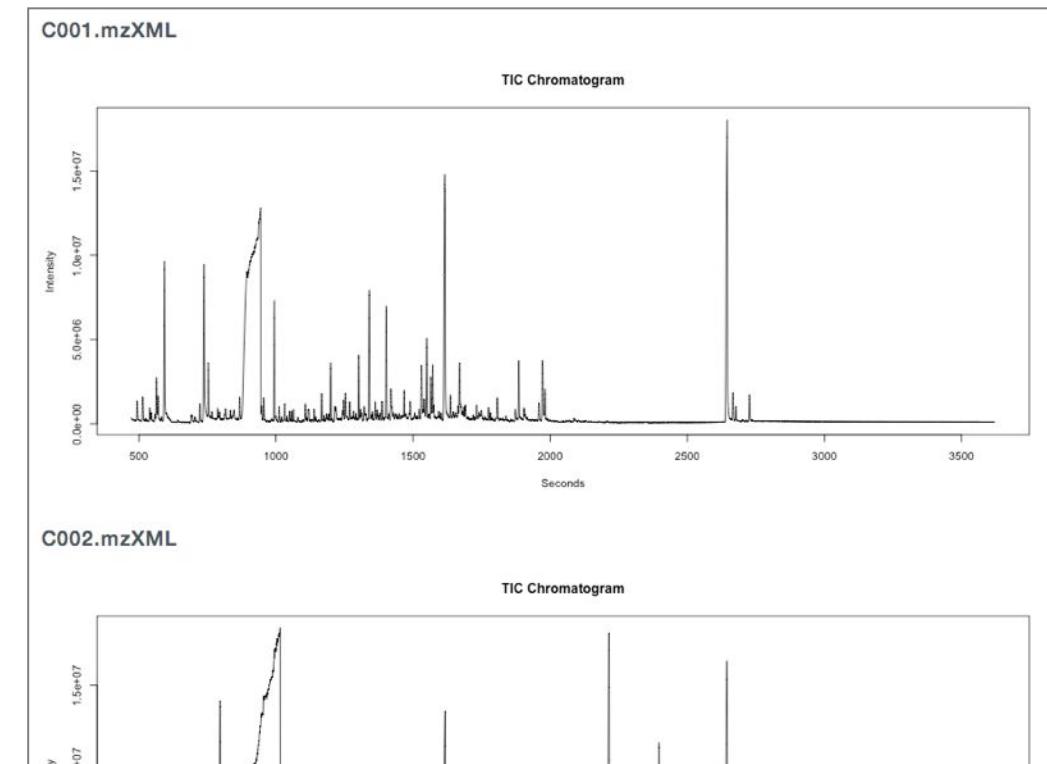
OR, run one of our examples:

Check Alkane Standard & Each Sample Spectrum

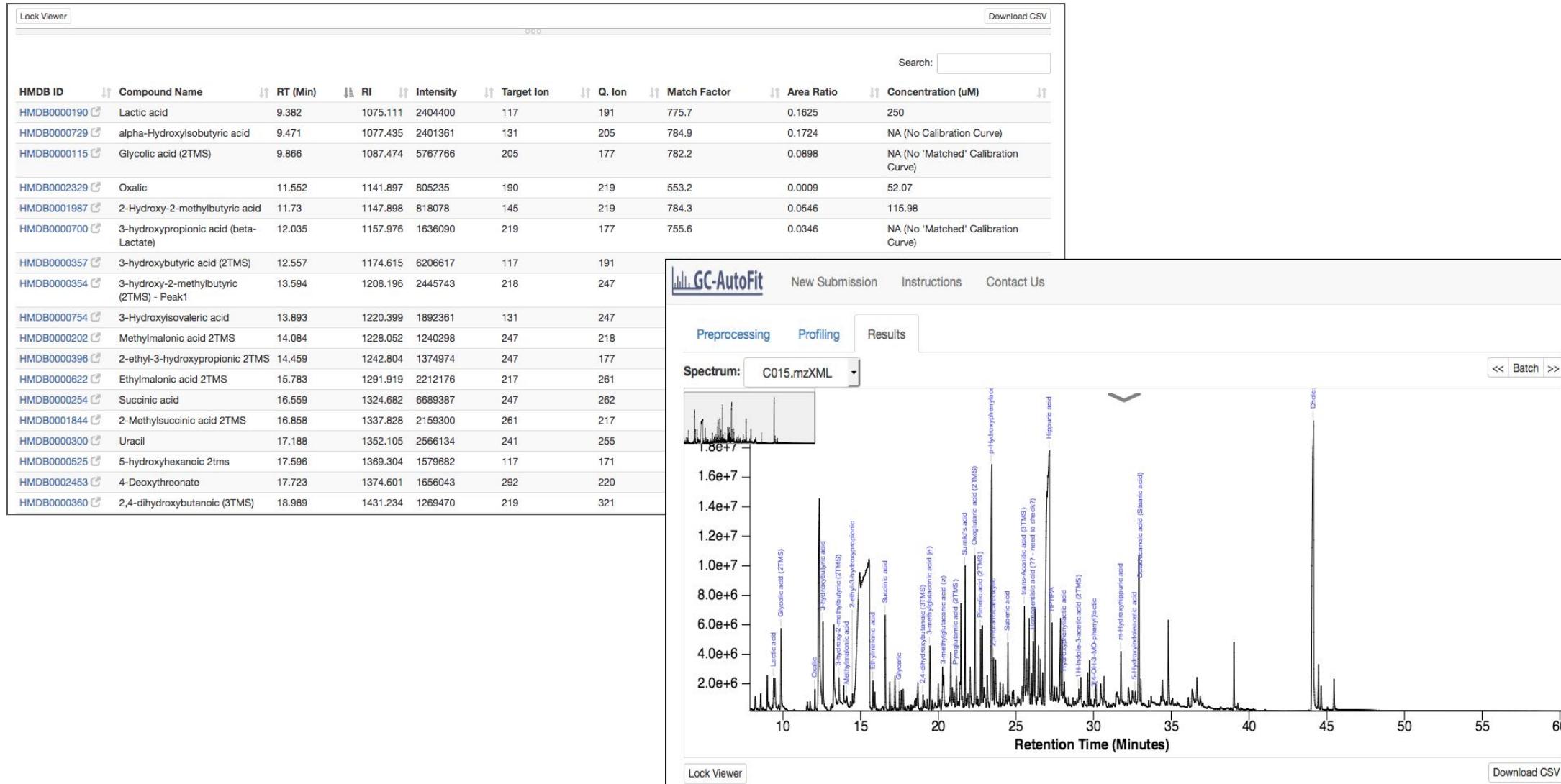
Alkane Standard Peaks and Blank Spectrum



Sample Spectrum

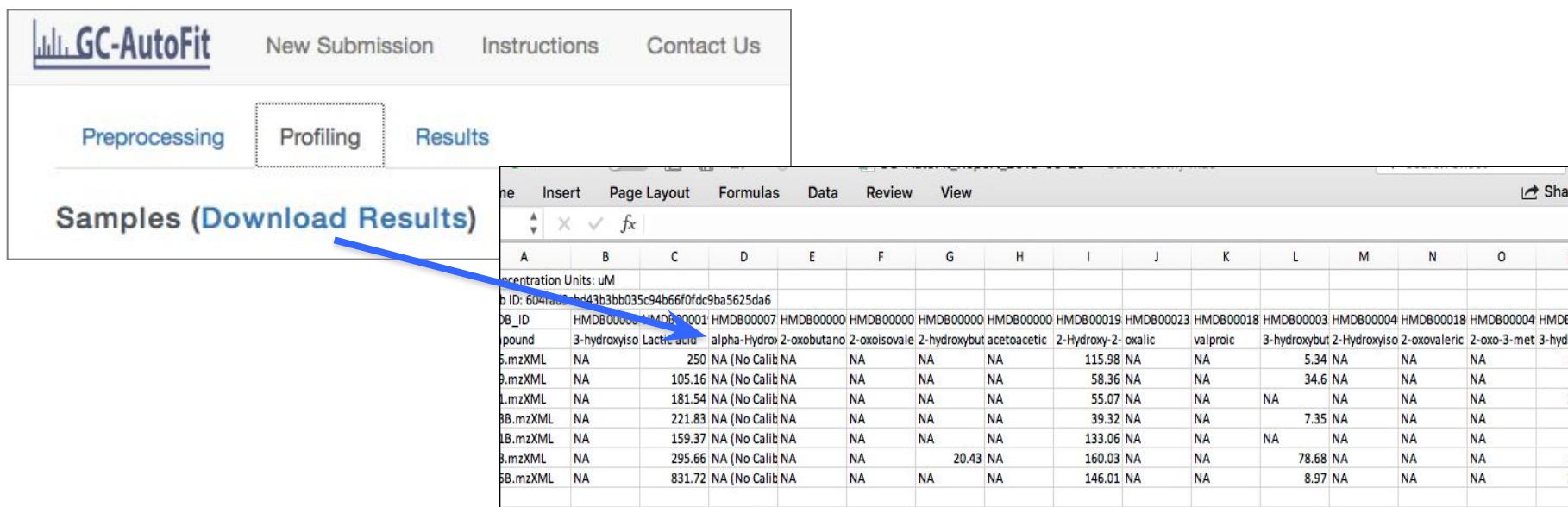


Profiling and Quantification



GC-AutoFit Final Results

- Table (CSV format file)
 - Table for each sample
 - Merged concentration for all samples
- Spectrum Viewer
 - Spectra with assigned compound names



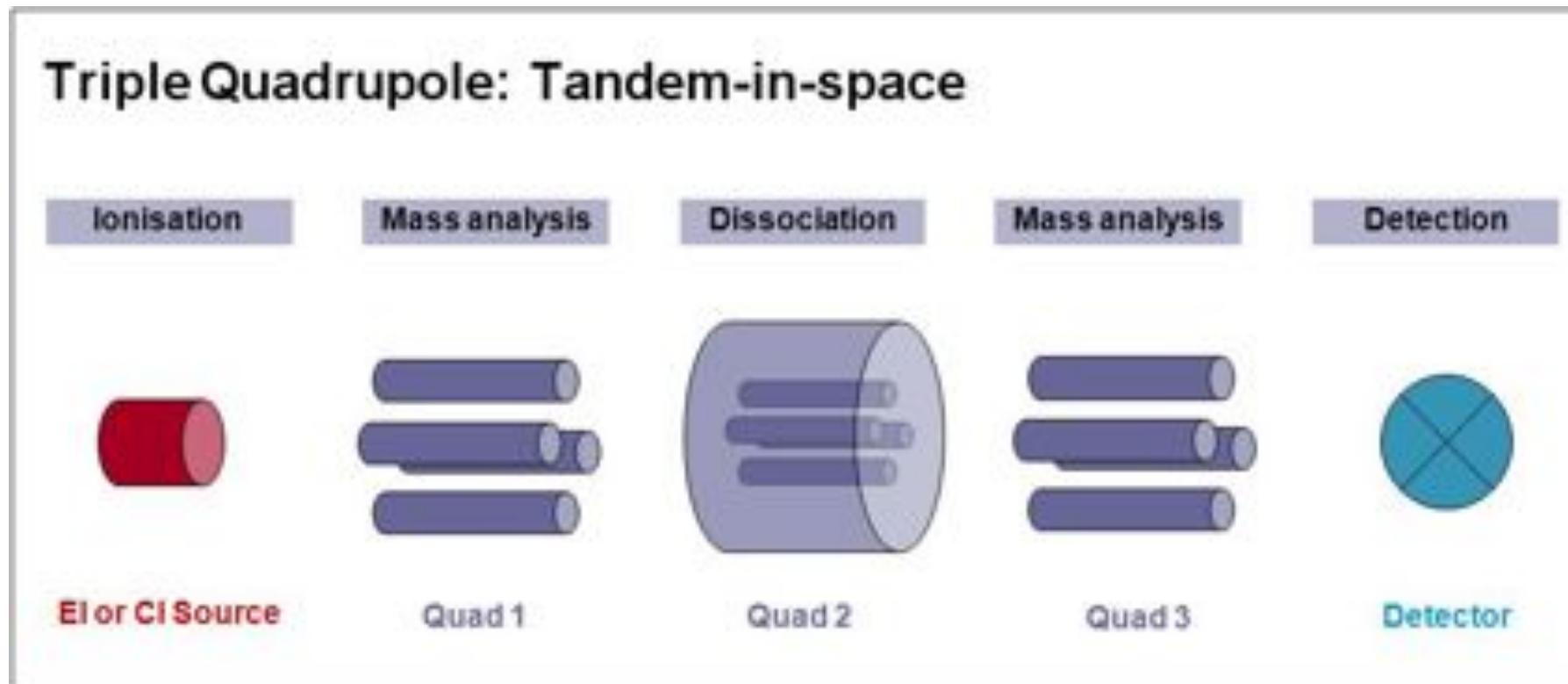
The screenshot shows the GC-AutoFit software interface. At the top, there is a navigation bar with links for "New Submission", "Instructions", and "Contact Us". Below the navigation bar, there are three tabs: "Preprocessing", "Profiling" (which is selected), and "Results". A blue arrow points from the text "Samples (Download Results)" to the "Results" tab. The main area displays a CSV table titled "Samples (Download Results)". The table has columns labeled A through P. The first few rows of data are as follows:

A	B	C	D	E	F	G	H	I	J	K	L	M	N	O	P
Concentration Units: uM															
Sample ID:	HMDB000001	HMDB000001	HMDB000007	HMDB000000	HMDB000000	HMDB000000	HMDB000019	HMDB00023	HMDB00018	HMDB00003	HMDB00004	HMDB00018	HMDB00004	HMDB00004	HMDB00004
Compound	3-hydroxyiso	Lactic acid	alpha-Hydrox	2-oxobutano	2-oxoisoval	2-hydroxybut	acetooacetic	2-Hydroxy-2-	oxalic	valproic	3-hydroxybut	2-Hydroxyiso	2-oxovaleric	2-oxo-3-met	3-hyd
5.mzXML	NA	250	NA (No Calib)	NA	NA	NA	115.98	NA	NA	5.34	NA	NA	NA	NA	1
9.mzXML	NA	105.16	NA (No Calib)	NA	NA	NA	58.36	NA	NA	34.6	NA	NA	NA	NA	
1.mzXML	NA	181.54	NA (No Calib)	NA	NA	NA	55.07	NA	NA	NA	NA	NA	NA	NA	2
8B.mzXML	NA	221.83	NA (No Calib)	NA	NA	NA	39.32	NA	NA	7.35	NA	NA	NA	NA	
1B.mzXML	NA	159.37	NA (No Calib)	NA	NA	NA	133.06	NA	NA	NA	NA	NA	NA	NA	
8.mzXML	NA	295.66	NA (No Calib)	NA	NA	20.43	NA	160.03	NA	NA	78.68	NA	NA	NA	1
5B.mzXML	NA	831.72	NA (No Calib)	NA	NA	NA	146.01	NA	NA	8.97	NA	NA	NA	NA	2

LC-MS Metabolomics

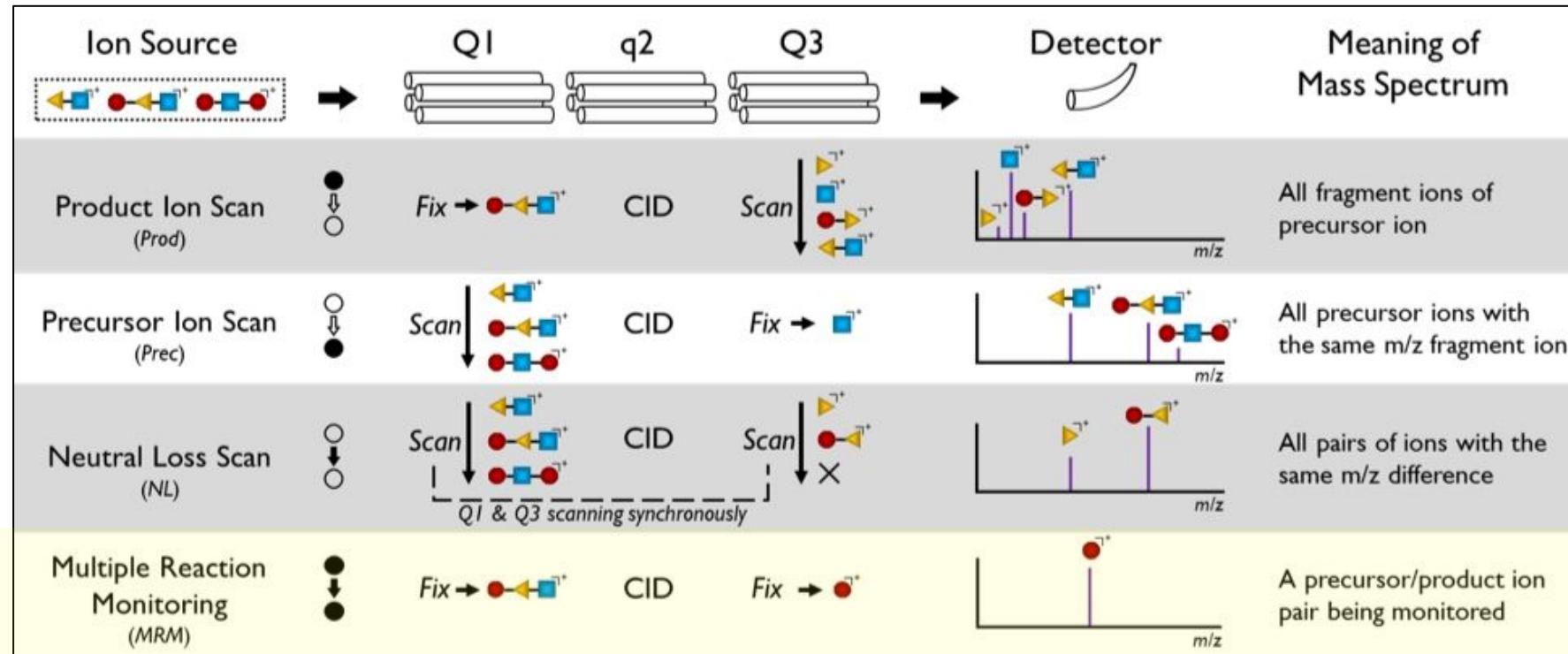
Targeted Metabolomics by LC-MS/MS

- Typically requires the use of a QqQ (or tandem) mass spectrometer to perform selective mass filtering



Targeted Metabolomics by LC-MS/MS

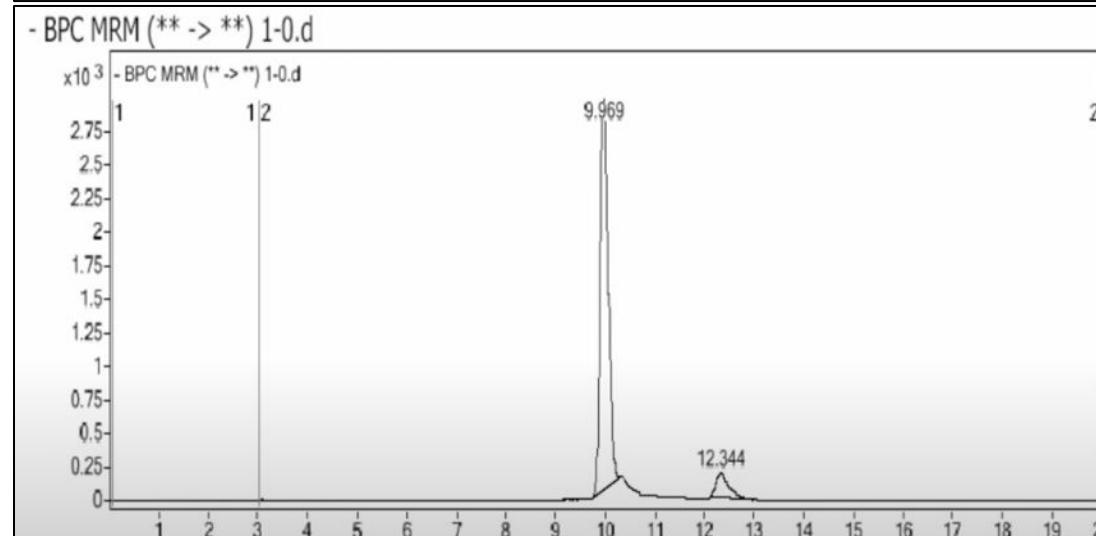
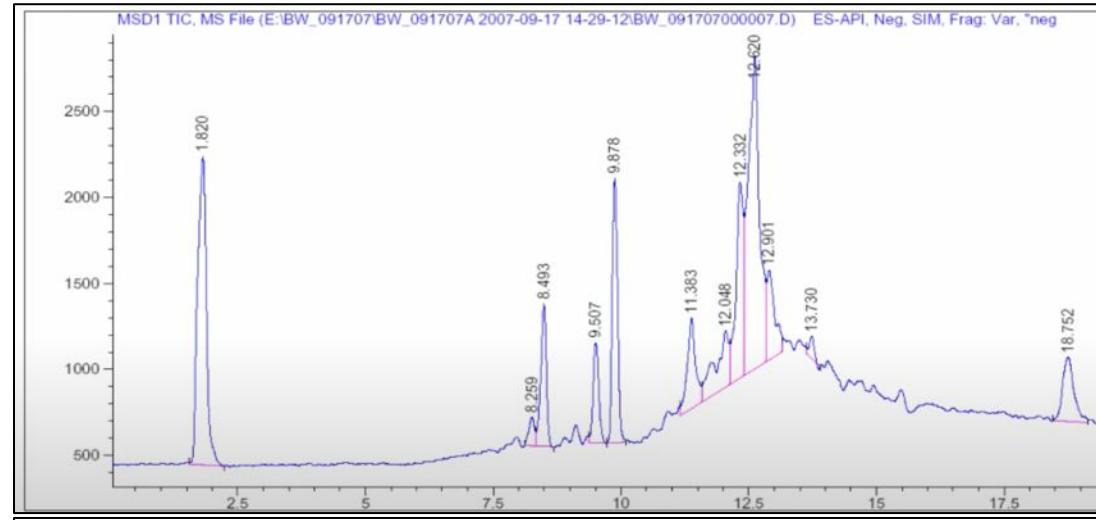
- Selective mass filtering involves choosing **precursor + product ion pairs** and specific retention times that correspond to specific known molecules - called MRM



MRMs provide a means of sorting through extraneous material

TIC (Single Quad) vs TIC (QqQ-MRM) of Serum

- **Top Spectrum – Single quad TIC showing overlapped peaks (hard to integrate, hard to quantify)**
- **Bottom Spectrum – QqQ with MRM choosing 286/152 m/z pair that elutes at 9.89 minutes (clean, easy to integrate and quantify)**



Targeted Metabolomics by LC-MS/MS – Step 1

- A list of precursor + product ion pairs (Q1/Q3), collision energies, declustering potentials and RTs is determined for all the target molecules of interest and used by the MS instrument
- These are also used by the analysis software to ID MRM peaks, perform integration and perform ISTD calibrations

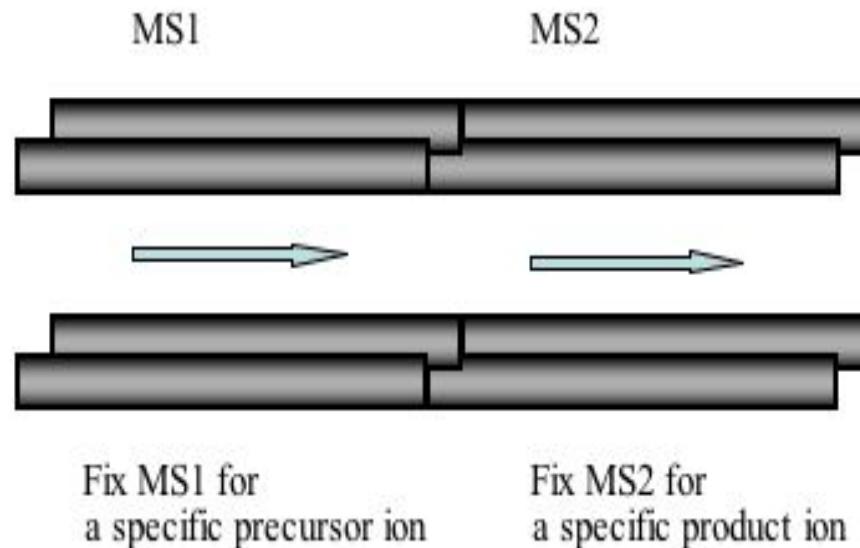


Table I +MRM Transitions for Opiates.

Mass Spectrometer Experiments:

Compound	Q1	Q3	Declustering Potential (V)	Collision Energy (V)
Morphine	286	152	46	79
Morphine	286	165	46	51
Hydromorphone	286	185	46	41
Hydromorphone	286	157	46	55
Oxymorphone	302	227	36	37
Oxymorphone	302	198	36	55
Codeline	300	152	46	85
Codeline	300	115	46	89
Hydrocodone	300	199	46	39
Hydrocodone	300	128	46	39
Oxycodone	316	240	31	39
Oxycodone	316	256	31	33
6-Monoacetylmorphine	328	211	51	55
6-Monoacetylmorphine	328	193	51	35

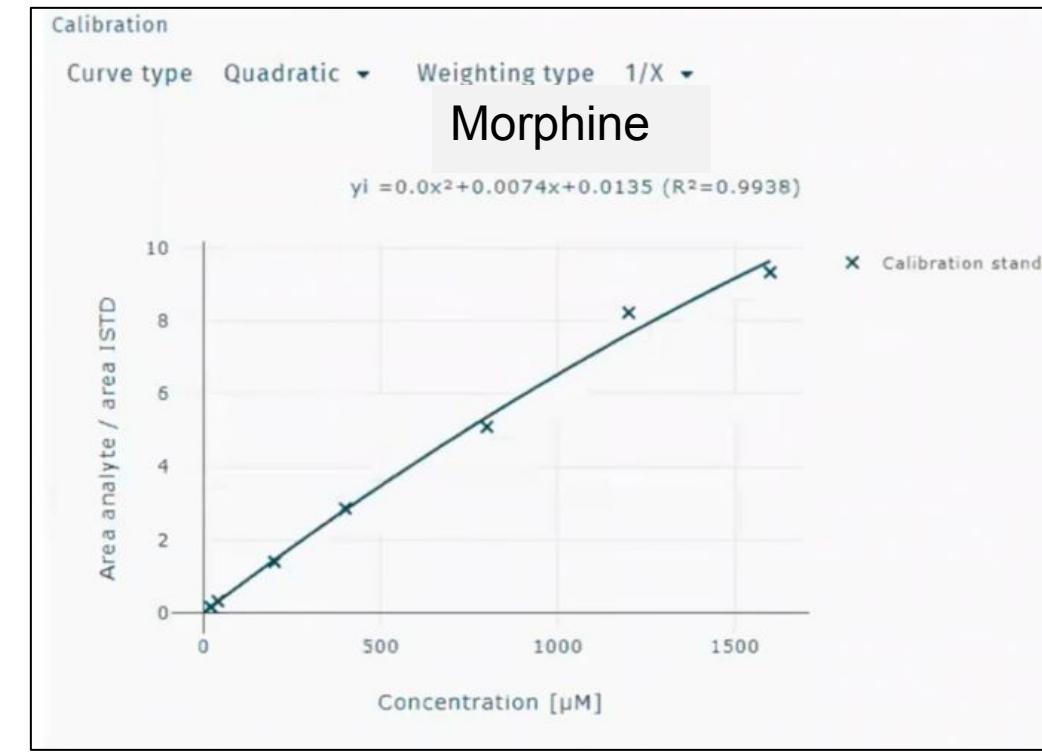
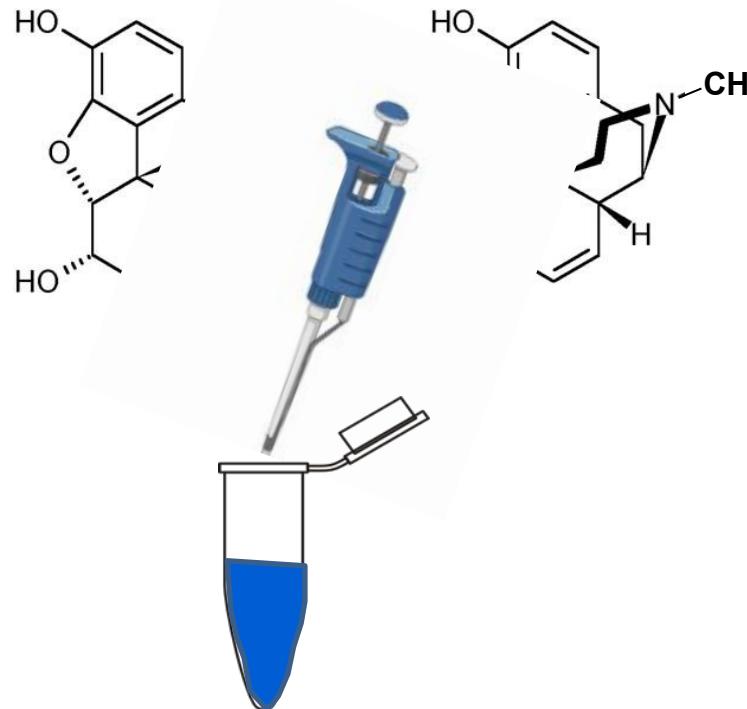
Targeted Metabolomics by LC-MS/MS – Step 2

- Samples and sample types are “registered” by the software to distinguish samples and MS spectra from calibrants and samples from blanks or other standards

A1	A2	A3	A4						
B1	B2	B3		B5					
C1	C2	35	C4						
D1	D2	D3	D4						
E1	E2	E3	E4						
F1	F2	F3	F4						
G1	G2	G3	G4						
H1	H2	H3							

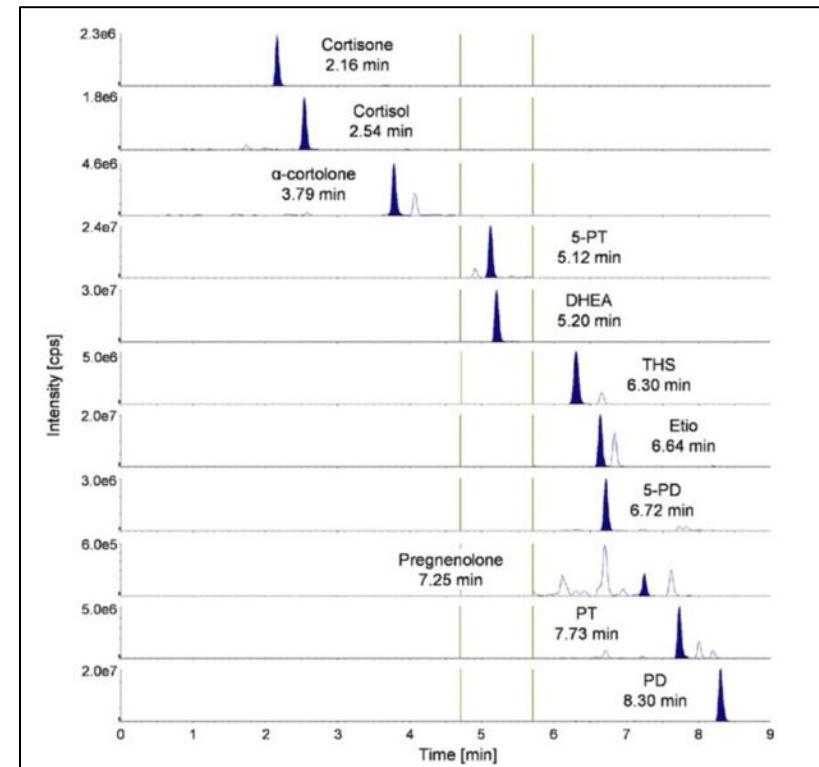
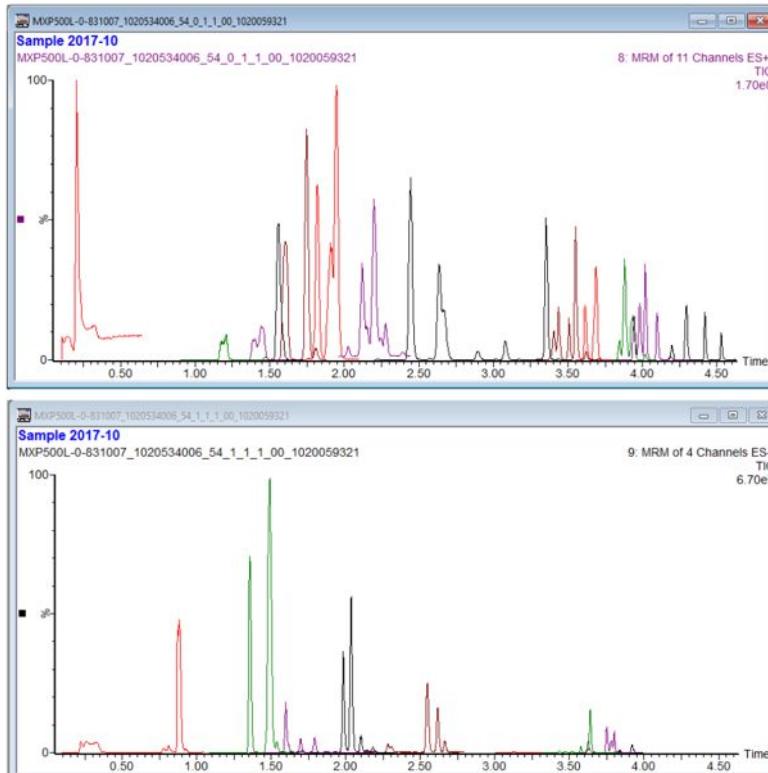
Targeted Metabolomics by LC-MS/MS – Step 3

- Internal and external isotopic standards (ISTDs) are analyzed and calibration curves are generated to help with integration and quantification (using the MRM list)



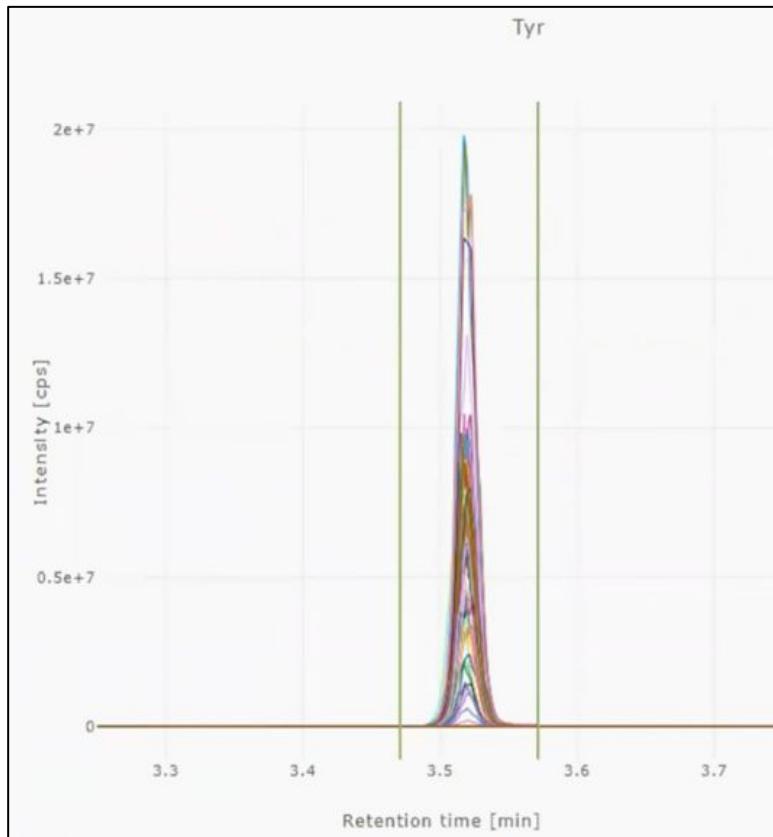
Targeted Metabolomics by LC-MS/MS – Step 4

- MRM transitions are identified from biological samples, qualifier ions are identified, peaks are automatically picked and all picked peaks are integrated to determine concentrations



Targeted Metabolomics by LC-MS/MS – Step 5

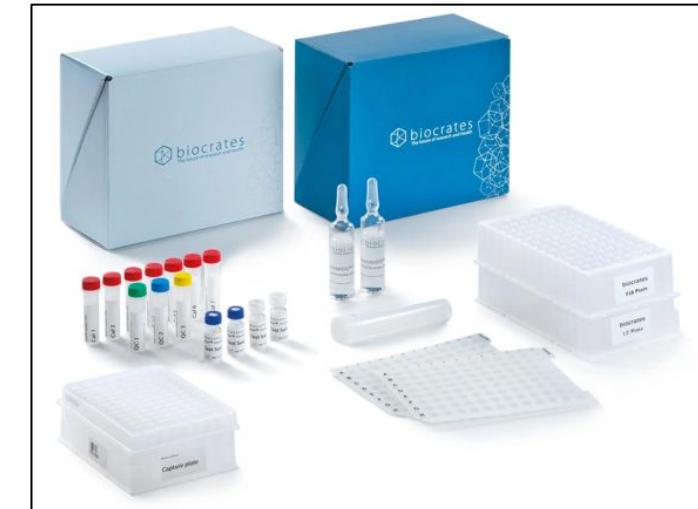
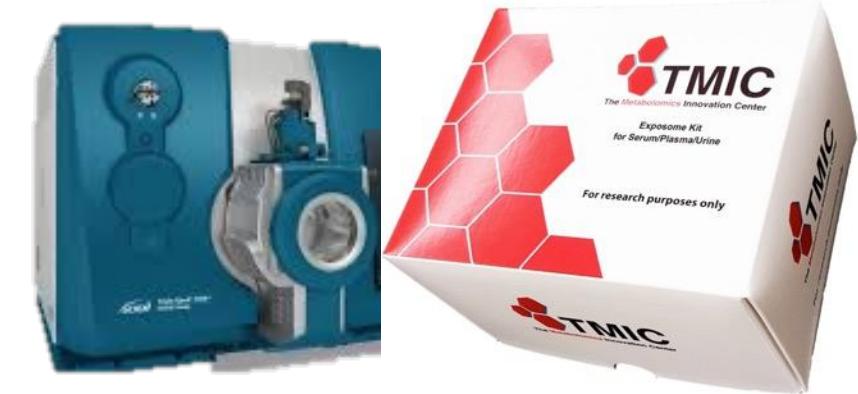
- Integrated areas and concentrations across all metabolites are determined and a list of metabolite concentrations is generated



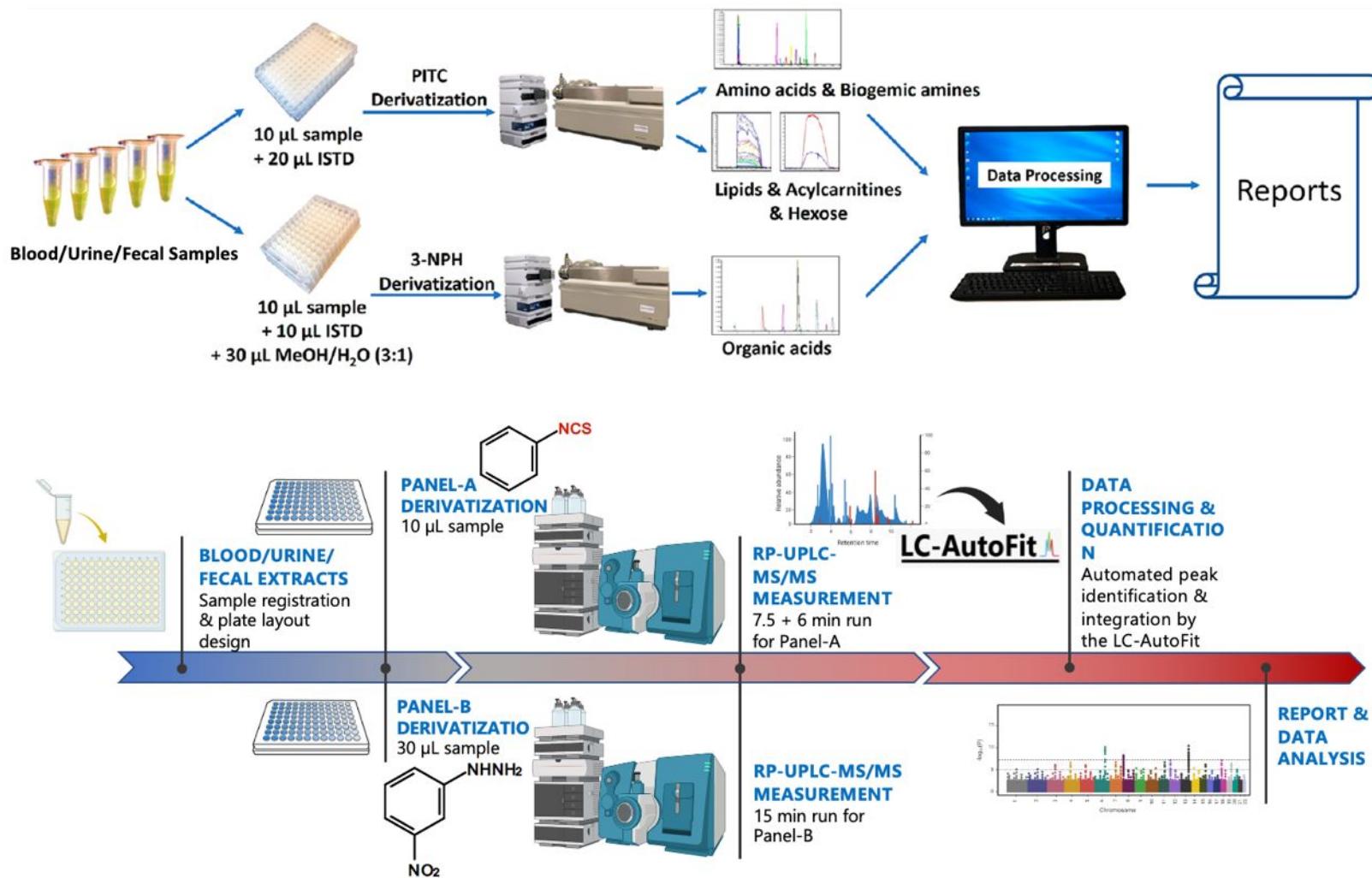
Compound	Retention Time (min)	Conc. in Urine (μM)	Compound	Retention Time (min)	Conc. in Urine (μM)
Dns-o-phospho -L-serine	0.92	<D.L.*	Dns-Ile	6.35	25
Dns-o-phospho -L-tyrosine	0.95	<D.L.	Dns-3-aminosalicylic acid	6.44	0.5
Dns-adenosine monophosphate	0.99	<D.L.	Dns-pipeolic acid	6.50	0.5
Dns-o-phosphoethanolamine	1.06	16	Dns-Leu	6.54	54
Dns-glucosamine	1.06	22	Dns-cystathione	6.54	0.3
Dns-o-phospho -L-threonine	1.09	<D.L.	Dns-Leu-Pro	6.60	0.4
Dns-6-dimethylamine putine	1.20	<D.L.	Dns-5-hydroxylysine	6.65	1.6
Dns-3-methyl -histidine	1.22	80	Dns-Cysteine	6.73	160
Dns-taurine	1.25	834	Dns-N-norleucine	6.81	0.1
Dns-carnosine	1.34	28	Dns-5-hydroxyproline	7.17	<D.L.
Dns-Arg	1.53	36	Dns-dimethylamine	7.33	293
Dns-Asn	1.55	133	Dns-5-HIAA	7.46	18
Dns-hypotaurine	1.58	10	Dns-umbelliferon	7.47	1.9
Dns-homocarnosine	1.61	3.9	Dns-2,3-diaminopropionic acid	7.63	<D.L.
Dns-guanidine	1.62	<D.L.	Dns-L-ornithine	7.70	15
Dns-Gln	1.72	633	Dns-4-acetylaminophenol	7.73	51
Dns-allantoin	1.83	3.8	Dns-procaine	7.73	8.9
Dns-L-citulline	1.87	2.9	Dns-homocystine	7.76	3.3
Dns-1 (or 3-) methylhistamine	1.94	1.9	Dns-acetaminophen	7.97	82
Dns-adenosine	2.06	2.6	Dns-Phe-Phe	8.03	0.4
Dns-methylguanidine	2.20	<D.L.	Dns-5-methoxy salicylic acid	8.04	2.1
Dns-Ser	2.24	511	Dns-Lys	8.16	184
Dns-aspartic acid amide	2.44	26	Dns-aniline	8.17	<D.L.
Dns-4-hydroxy -proline	2.56	2.3	Dns-leu-Phe	8.22	0.3
Dns-Glu	2.57	21	Dns-His	8.35	1550
Dns-Asp	2.60	90	Dns-4-thiolsine	8.37	<D.L.
Dns-Thr	3.03	157	Dns-benzylamine	8.38	<D.L.
Dns-epinephrine	3.05	<D.L.	Dns-1-ephedrine	8.50	0.6
Dns-ethanolamine	3.11	471	Dns-tryptamine	8.63	0.4
Dns-aminoacidic acid	3.17	70	Dns-pyridoxamine	8.94	<D.L.
Dns-Gly	3.43	2510	Dns-2-methyl -benzylamine	9.24	<D.L.
Dns-Ala	3.88	593	Dns-5-hydroxytryptophan	9.25	0.12
Dns-aminolevulinic acid	3.97	30	Dns-1,3-diaminopropane	9.44	0.23
Dns-o-amino -butyric acid	3.98	4.6	Dns-purinesine	9.60	0.5
Dns-p-amino -hippuric acid	3.98	2.9	Dns-1,2-diaminopropane	9.66	0.1
Dns-5-hydroxymethyluricil	4.58	1.9	Dns-tyrosinamide	9.79	29
Dns-tryptophanamide	4.70	5.5	Dns-dopamine	10.08	140
Dns-isoguanine	4.75	<D.L.	Dns-cadaverine	10.08	0.08
Dns-5-aminopentanoic acid	4.79	1.6	Dns-histamine	10.19	0.4
Dns-sarcosine	4.81	7.2	Dns-3-methoxy -tyramine	10.19	9.2
Dns-3-amino -isobutyrate	4.81	85	Dns-Tyr	10.28	321
Dns-2-aminobutyric acid	4.91	17	Dns-cysteamine	10.44	<D.L.

LC-MS Kits

- Quantitative, kit-based, targeted MS metabolomics systems with hundreds of chemical standards, bundled software and ISO-level protocols
- Permit 140-640 compounds (soon 1200+) to be accurately measured
- Easy to use, training videos and ported to several labs already
- **\$45-\$90/sample, 96 well format**



LC-MS Kit - Overview



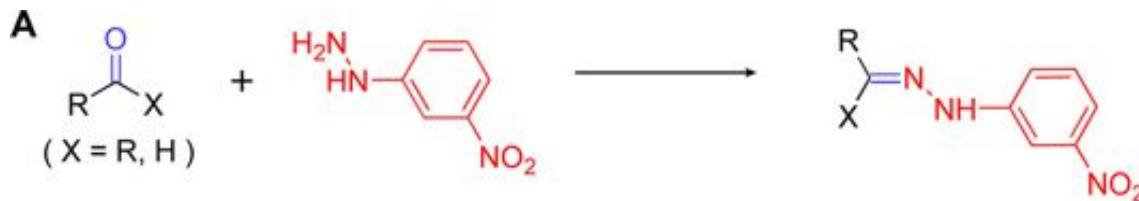
Design Considerations

- 96 well format (for automation)
- Mix of LC-MS and FIA (flow injection analysis)
- <50 uL of material required
- <30 minutes instrument time per sample
- Needs to cover: organic acids, amino acids, biogenic amines, indoles, sugars, acylcarnitines, lipids (PCs, LPCs, SMs, TGs, DGs, CEs)
- Needs to have own web-based software to avoid compatibility and portability issues
- Data processing time needs to be <2 min sample

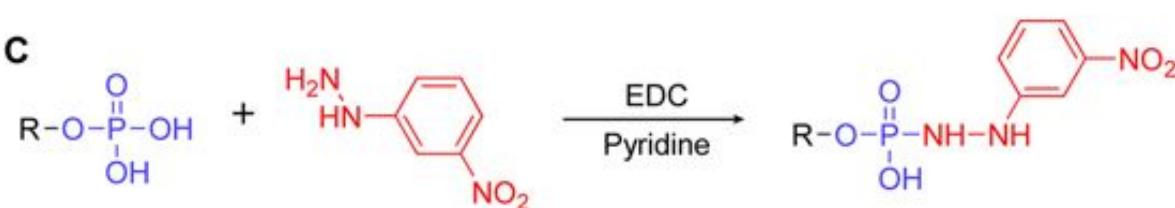
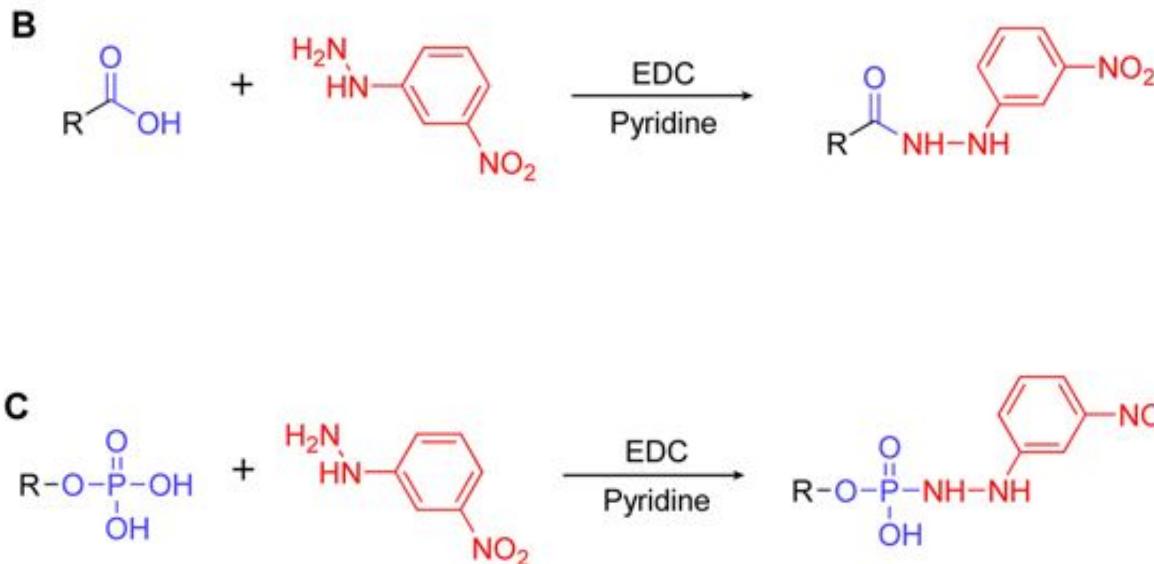
Design Considerations

- Use chemical tagging to allow greater use of RP-HPLC
- Two panels (one for amines, one for organic acids) + FIA for lipids and acylcarnitines
- Isotopic standards (ISTDs) for all panels and FIA runs
- Inter-day QC accuracy and recovery $100 \pm 10\%$
- Inter-day QC RSD accuracy and recovery <10%
- CV <15%, LOD < 1 uM on >95%, LOQ < 1 uM on >90% of measured compounds

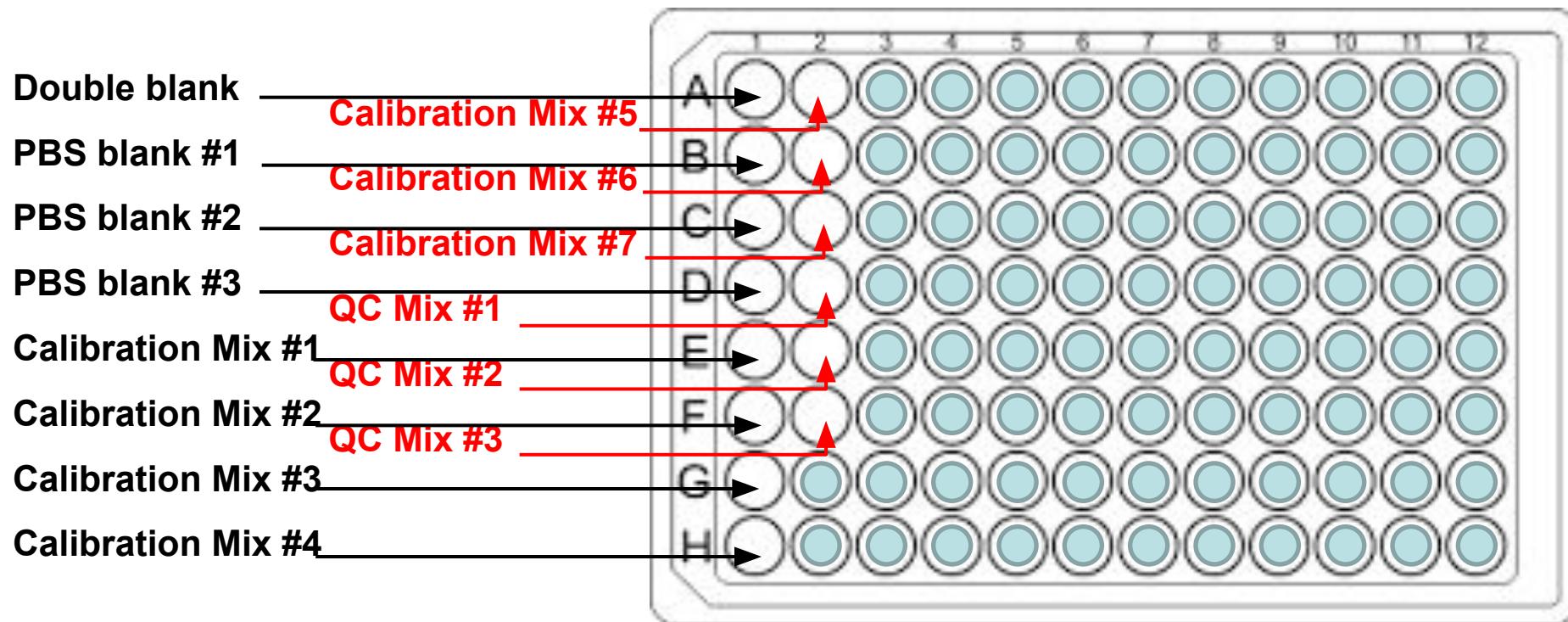
Chemical Tagging



- **^{13}C labeled 3-NPH (3-nitrophenylhydrazine) for organic acids**
- **Improves chemical stability**
- **Increases hydrophobicity and improves RPLC separation**
- **Improves mass spectrometry ionization/detection**
- **Reduces cost (avoids individual isotope-labeled standards)**



96 Well Plate Design



**14 Standards/Blanks/Calibrations
+82 Samples**

Three Assay Types

Big

Bigger

Bigest

TMIC PRIME

Chemical Families	Number of metabolites
Amino acids and derivatives	42
Acylcarnitines	40
Organic acids	15
Lysophosphatidylcholines	14
Phosphatidylcholines	10
Sphingomyelins	10
Biogenic Amines	9
Ketone and Keto acids	2
Vitamins and Derivatives	1
Nucleoside/Nucleotide	1
Dipeptide	1
Total	145

+40 metabolite ratios

Grand Total = 185

TMIC MEGA

Chemical families	Number of Metabolites
Amino acids and derivatives	72
Organic acids	58
Nucleobases and Nucleosides	24
Biogenic Amines	23
Ketone and Keto acids	11
Indole derivatives	9
Vitamins and Derivatives	3
Sulfates	2
Others	2
Dipeptides	1
Triglycerides	212
Phosphatidylcholines	75
Acylcarnitines	40
Cholesteryl Esters	22
Diglycerides	21
Ceramides	18
Hexosylceramides	17
Lysophosphatidylcholines	14
Nucleobases and Nucleosides	32
Organic Acids	106
Others	14
Phosphatidylcholines	75
Sphingomyelins	14
Steroids and Hormones	11
Sugars, Sugar acids and Sugar alcohols	29
Sulfates	7
Triglycerides	445
Trihexosylceramides	4
Vitamins and Derivatives	5
Total	645

+250 metabolite ratios

Grand Total = 895

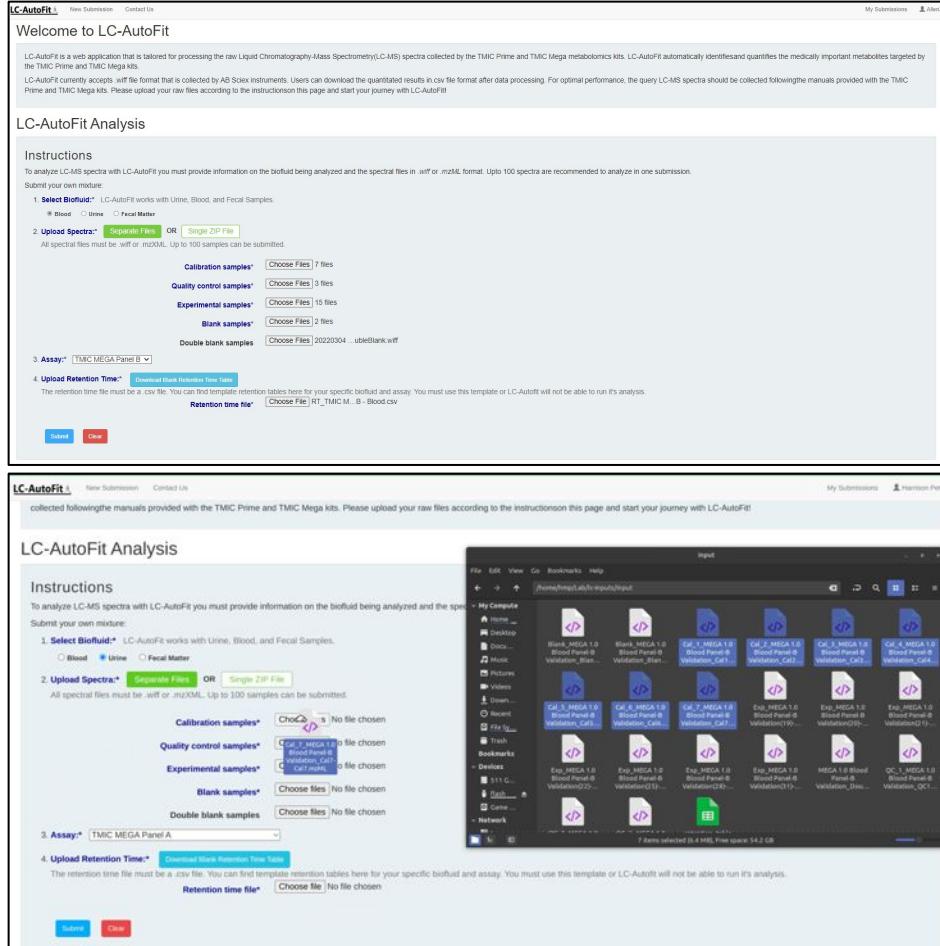
TMIC GIGA

Chemical families	Number of Metabolites
Acylcarnitines	86
Aldehydes	6
Amino Acids and Derivatives	103
Bile Acids	20
Biogenic Amines	26
Ceramides	14
Cholesteryl Esters	22
Diglycerides	190
Dihexosylceramides	7
Dihydroceramides	4
Dipeptides	52
Fatty Acids and Derivatives	25
Hexosylceramides	17
Indole Derivatives	11
Ketone and Keto acids	20
Lysophosphatidylcholines	14
Nucleobases and Nucleosides	32
Organic Acids	106
Others	14
Phosphatidylcholines	75
Sphingomyelins	14
Steroids and Hormones	11
Sugars, Sugar acids and Sugar alcohols	29
Sulfates	7
Triglycerides	445
Trihexosylceramides	4
Vitamins and Derivatives	5
Total	1359

+450 metabolite ratios

Grand Total = 1800

LC-AutoFit (Automated LC-MS)



Welcome to LC-AutoFit

LC-AutoFit is a web application that is tailored for processing the raw Liquid Chromatography-Mass Spectrometry(LC-MS) spectra collected by the TMC Prime and TMC Mega metabolomics kits. LC-AutoFit automatically identifies and quantifies the medically important metabolites targeted by the TMC Prime and TMC Mega kits.

LC-AutoFit currently accepts .wif file format that is collected by AB Sciex instruments. Users can download the quantitated results in csv file format after data processing. For optimal performance, the query LC-MS spectra should be collected following the manuals provided with the TMC Prime and TMC Mega kits. Please upload your raw files according to the instructions on this page and start your journey with LC-AutoFit!

LC-AutoFit Analysis

Instructions

To analyze LC-MS spectra with LC-AutoFit you must provide information on the biofluid being analyzed and the spectral files in .wif or .mzXML format. Up to 100 spectra are recommended to analyze in one submission.

Submit your own mixture:

1. Select Biofluid: Urine, Blood, Fecal Matter

2. Upload Spectra: Separate Files OR Single ZIP File

All spectral files must be .wif or .mzXML. Up to 100 samples can be submitted.

Calibration samples: Choose Files 7 files

Quality control samples: Choose Files 3 files

Experimental samples: Choose Files 15 files

Blank samples: Choose Files 2 files

Double blank samples: Choose Files 20220304_uebeBlank.wif

3. Assay: TMC MEGA Panel A

4. Upload Retention Time: Download Blank Retention Time Table

The retention time file must be a .csv file. You can find template retention tables here for your specific biofluid and assay. You must use this template or LC-AutoFit will not be able to run its analysis.

Retention time file: Choose File RT_TMC M... - B - Blood.csv

LC-AutoFit Analysis

Instructions

To analyze LC-MS spectra with LC-AutoFit you must provide information on the biofluid being analyzed and the spectral files in .wif or .mzXML format. Up to 100 samples can be submitted.

Submit your own mixture:

1. Select Biofluid: Urine, Blood, Fecal Matter

2. Upload Spectra: Separate Files OR Single ZIP File

All spectral files must be .wif or .mzXML. Up to 100 samples can be submitted.

Calibration samples: Choose Files No file chosen

Quality control samples: Choose Files No file chosen

Experimental samples: Choose Files No file chosen

Blank samples: Choose Files No file chosen

Double blank samples: Choose Files No file chosen

3. Assay: TMC MEGA Panel A

4. Upload Retention Time: Download Blank Retention Time Table

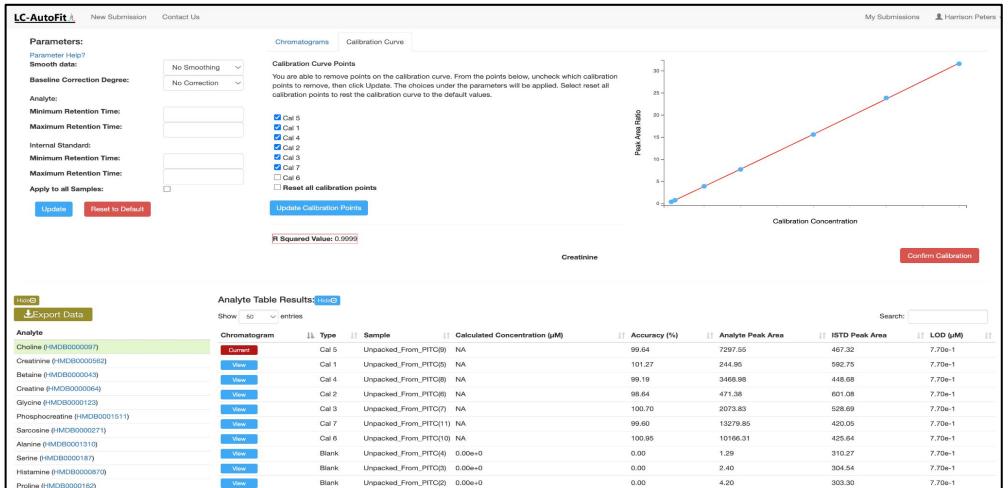
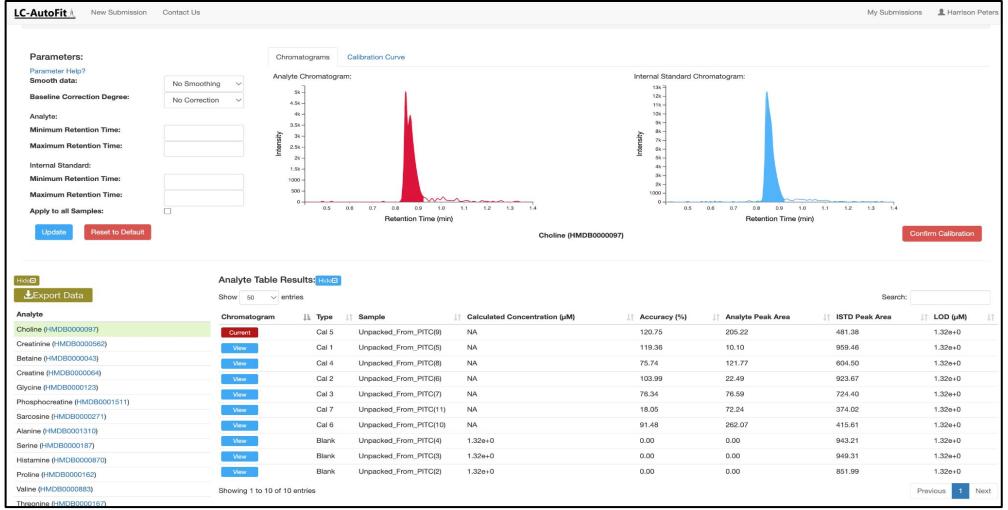
The retention time file must be a .csv file. You can find template retention tables here for your specific biofluid and assay. You must use this template or LC-AutoFit will not be able to run its analysis.

Retention time file: Choose File No file chosen

- Requires user to select biofluid
- Requires user to provide MS/MS spectral files and to indicate calibration files, sample files, QC sample files & blank sample files
- Retention time file and retention time list is also required
- Files can be dragged and dropped into LC-AutoFit browser to simplify file uploads
- Processes 96 samples in just 5-10 minutes
- 180-640 compounds can be identified and quantified
- Optimized for blood, urine, stool samples for Qtrap instruments

<http://dev-lcautofit.wishartlab.com/>

LC-AutoFit (Automated LC-MS)



- Available as both a web-based server or as a dockerized system for local installation
- Takes ~60 minutes to ID and quantify 640 metabolites
- Currently being adapted to work on Thermo Orbitrap instruments as well as Waters and Agilent QqQ instruments
- Supports interactive calibration curve editing, interactive peak adjustment and peak integration

LC-AutoFit Final Results

- Table (CSV format file)
 - Table for each sample
 - Merged concentration for all samples
- Spectrum Viewer
 - Spectra with assigned compound names

The screenshot shows the LC-AutoFit software interface. At the top, there is a navigation bar with links for "New Submission" and "Contact Us". Below this, there are three tabs: "Preprocessing", "Calibration Optimization", and "Results". The "Results" tab is highlighted with a red circle and a red arrow points from it down to a green "Export Data" button. The main area displays a table of analytical results. The table has columns for "Analytes" and "HMDB ID", followed by 11 concentration values for various compounds. The compounds listed are Xanthine, Uric acid, p-Cresol sulfate, 4-Ethylphenyl sulfate, Indoxyl sulfate, Lactic acid, 3-Aminoisobutyric acid, Dimethylglycine, 2-Hydroxybutyric acid, and 2-Hydroxyisobutyric acid.

Analytes	Xanthine	Uric acid	p-Cresol sulfate	4-Ethylphenyl sulfate	Indoxyl sulfate	Lactic acid	3-Aminoisobutyric acid	Dimethylglycine	2-Hydroxybutyric acid	2-Hydroxyisobutyric acid	
HMDB ID	HMDB0000292	HMDB0000289	HMDB0011635	HMDB0062551	HMDB0000682	HMDB0000190	HMDB0003911	HMDB0000092	HMDB0000008	HMDB0000729	
LOD	9.0479	3.0633	2.226	0.0781	1.5484	0.0756	0.0016	0.0291	0.093	0.9181	
Exp_Sample_10061.mzML	109.4569	3467.0722	<LOD		0.5218	491.4485	153.9728	36.7247	10.5262	7.8116	37.34
Exp_Sample_10062.mzML	6.0283	417.0739	<LOD		0.1864	13.3977	<LOD		0.6748	0.8857	3.5664
Exp_Sample_10063.mzML	70.509	5096.3241	<LOD		0.4361	288.6951	34.9684	25.3835	6.9827	6.3335	29.9054
Exp_Sample_10064.mzML	489.6818	<LOD	<LOD		5.1416	581.7722	241.6214	86.2706	14.8949	17.3065	67.4947
Exp_Sample_10065.mzML	137.0416	<LOD	<LOD		0.9539	212.4157	67.2221	168.2361	27.4974	13.1586	61.0689
Exp_Sample_10066.mzML	62.0224	4795.3946	<LOD		0.4382	375.6666	31.2301	28.44	20.6412	5.5295	26.7932
Exp_Sample_10067.mzML	85.8285	5058.0803	<LOD		0.2564	160.7675	177.959	<LOD	2.9743	7.447	32.9374

Conclusion

- Targeted, quantitative metabolomics is easier and faster to do than untargeted metabolomics
- Targeted metabolomics can be done on 3 different platforms (NMR, GC-MS & LC-MS)
- Each analytical platform requires different workflows and different software
- Each platform has its advantages and disadvantages
- We will try out some of these tools and webservers in the upcoming lab

We are on a Coffee/Lunch Break & Networking Session

Workshop Sponsors:



Canadian Centre for
Computational
Genomics



HPC4Health

