

# Managing complexity:

“How many platforms do we need for metabolomics?”

Professor 1, March 10\_2010

“...a more complete picture as to the changes in these pathways., i.e. **dNTP pools** , **AMP levels, cAMP. etc.** We’d like get a handle on all of those if possible.

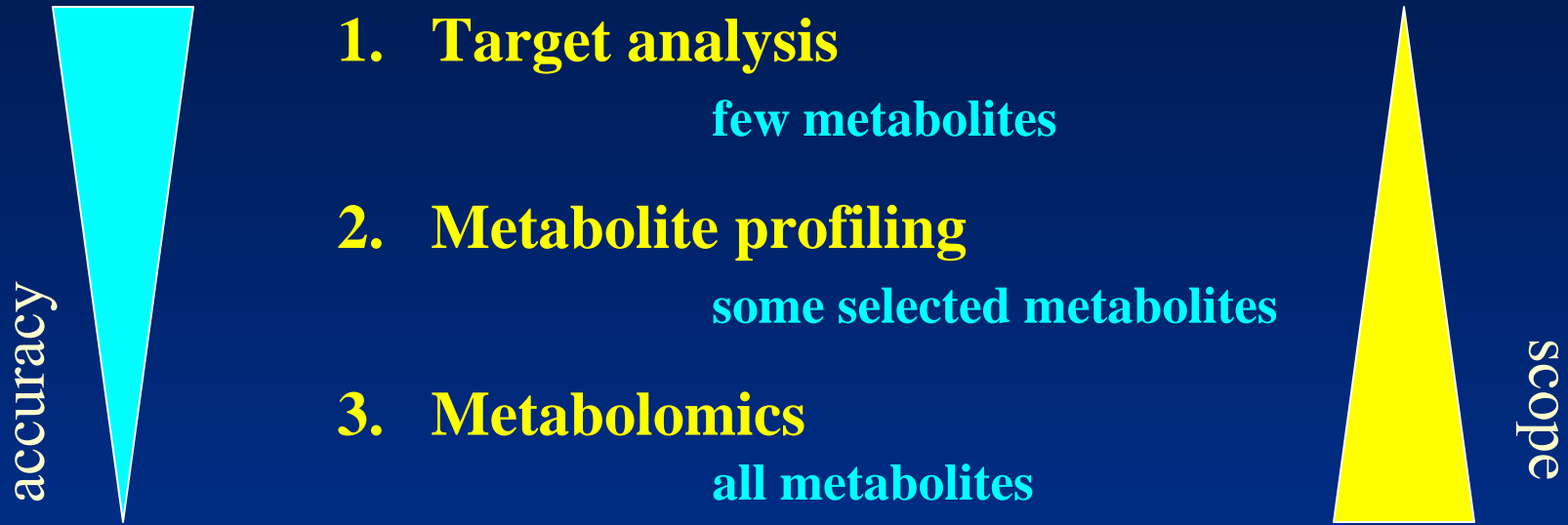
Professor 2, March 04\_2010

“...i.e. the levels of **MEP pathway metabolites, specifically 2C-methyl-D-erythritol 2,4-cylodiphosphate?**

Professor 3, March 01\_2010

“...in inflammation, where we believe the binding of this **specific oxylipid** to the receptor largely regulates the activation of the signaling cascade. Can you determine its structure?

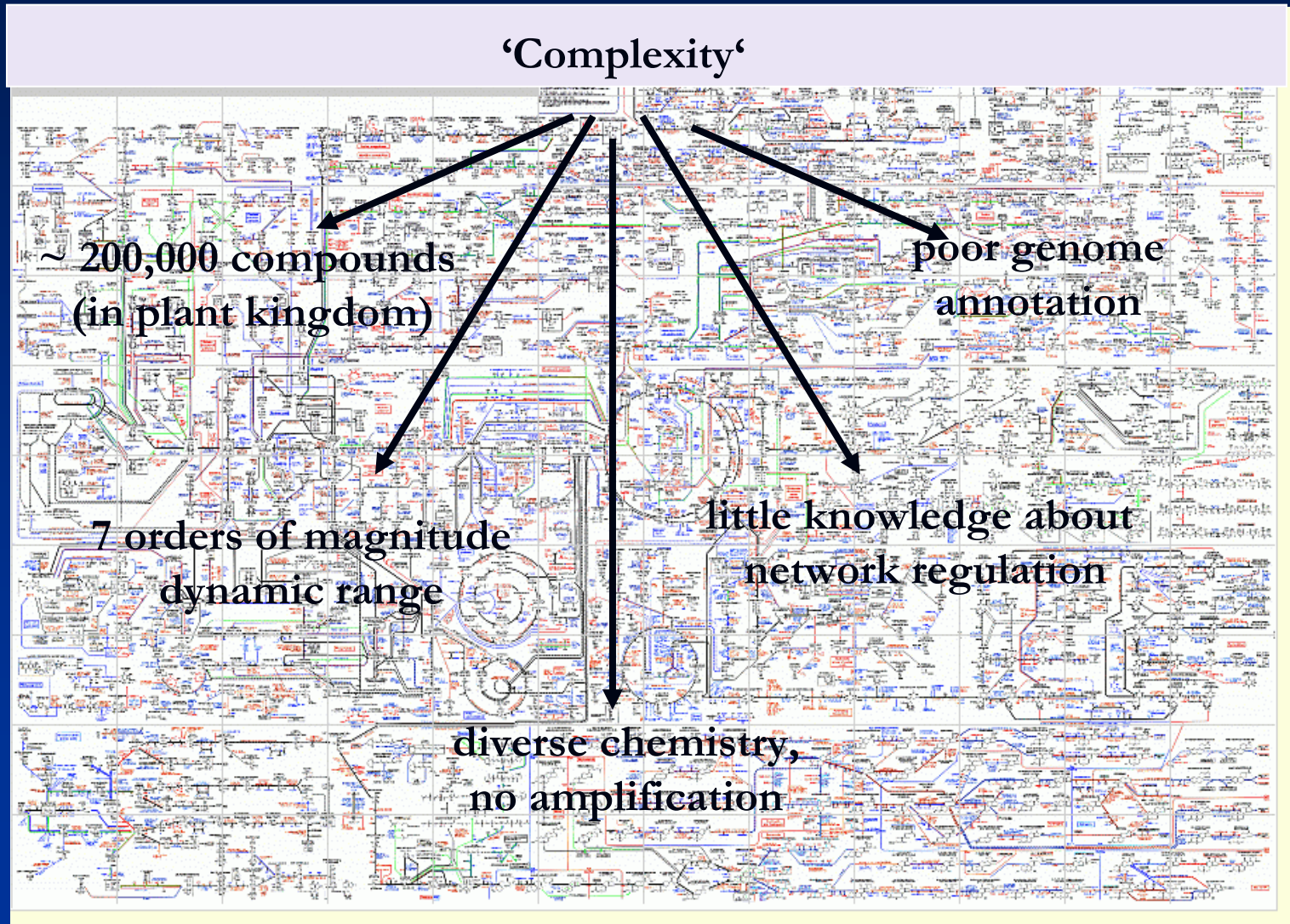
# What is metabolomics?



---

**Metabolic fingerprinting**  
classifying samples

# How can we separate, identify, quantify, store & compare thousands of metabolites?



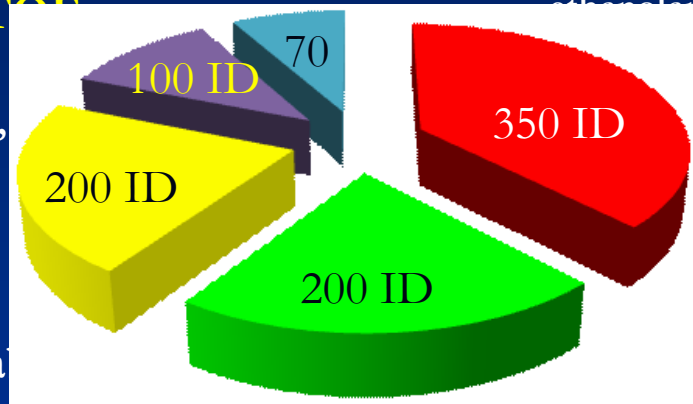
# 200,000+ known metabolites

## How many platforms do we need?

**pyGC-MS**  
 monomers  
 lignin, hemicellulose  
 complex lipids

**nanoESI-MS/MS** polar & neutral lipids  
**UPLC-MS/MS**  
 phosphatidylcholines, -serines,  
 sphingomyelins, plasmalogens, triglycerides

**Twister-GC-TOF**  
 volatiles  
 terpenes, alkanes,  
 FFA, benzenes



**GCxGC-TOF**  
 primary small meta  
 sugars, HO-acids, FFA, amino acids,  
 sterols, phosphates, aromatics

**UPLC-UV-MS/MS** secondary metabolites  
 oxylipids, anthocyanins, flavonoids, pigments  
 acylcarnitines, folates, glucuronidated & glycosylated aglycones



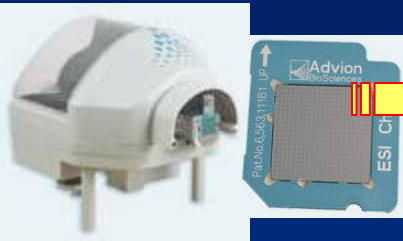
❖ Two or more independent parameters

# NanoESI-MS for membrane lipids

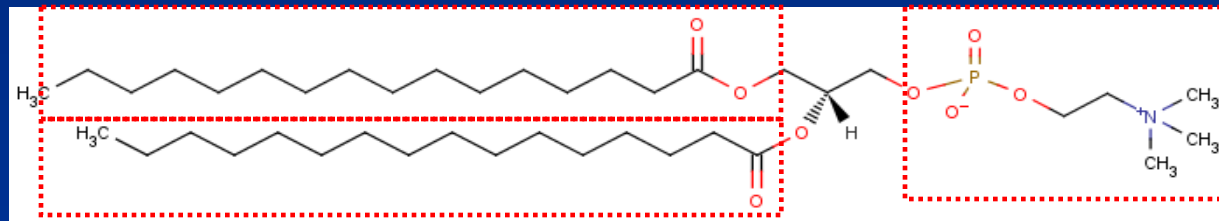
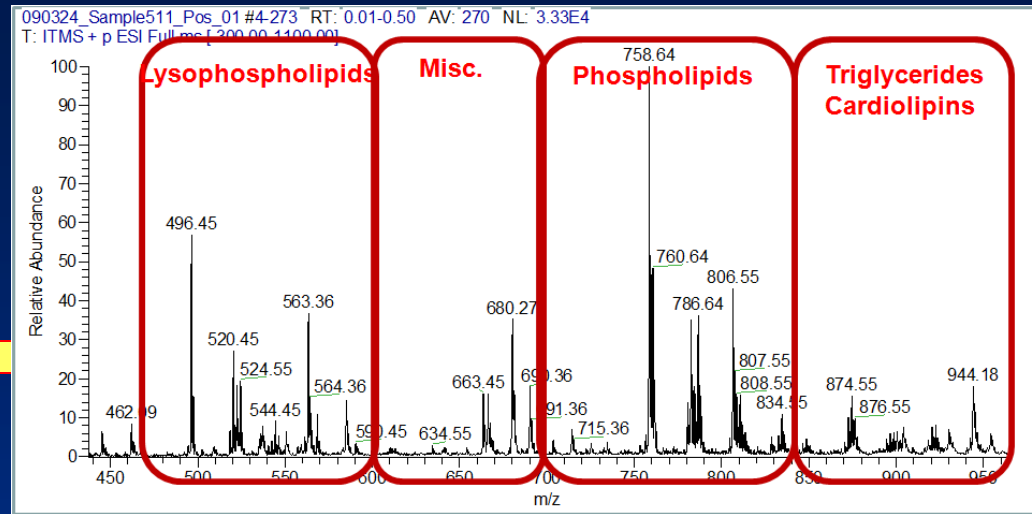
LTQ MS:  
Low resolution



NanoESI infusion  
Nanomate chip robot



LTQ-FT-ICR-MS  
High resolution

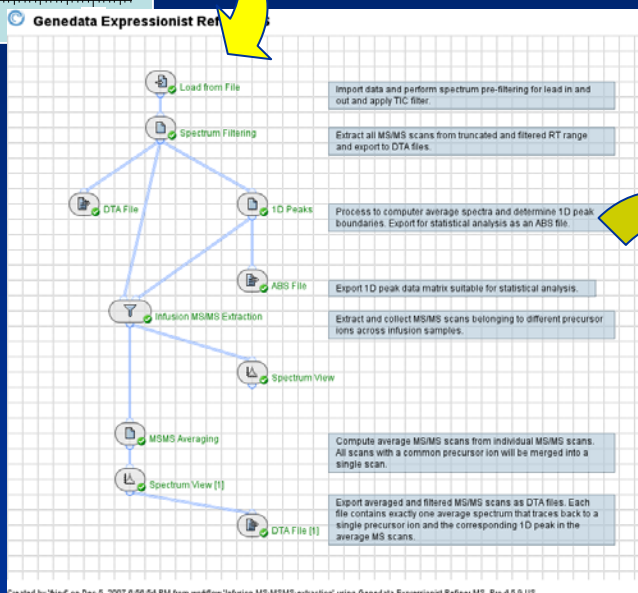
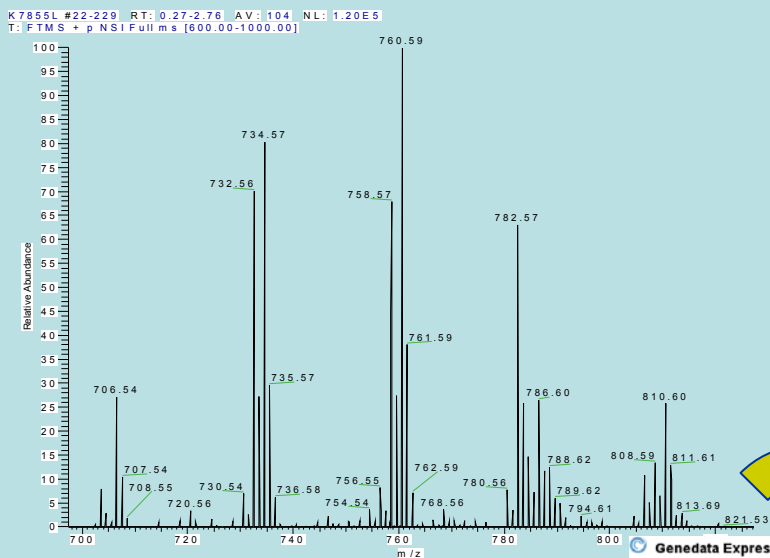


MS/MS fragmentation:

- head group
- sn1, sn2 acyl chains
- novel Fiehnlab LipidBLAST

# Lipidomics by nanoESI-ion trap MS

*blood plasma → up to 150 ID lipids*  
*Mostly abundant glycerophosphocholines.*  
*Data processing by Genedata Refiner MS*



RT	Abundance	m/z	Label	...
706.54	25	706.54		...
707.54	10	707.54		...
708.55	10	708.55		...
720.56	5	720.56		...
730.54	10	730.54		...
732.56	70	732.56		...
734.57	80	734.57		...
735.57	30	735.57		...
736.58	10	736.58		...
754.54	5	754.54		...
756.55	10	756.55		...
758.57	65	758.57		...
760.59	100	760.59		...
761.59	40	761.59		...
762.59	10	762.59		...
768.56	5	768.56		...
780.56	10	780.56		...
782.57	60	782.57		...
788.60	25	788.60		...
788.62	10	788.62		...
789.62	10	789.62		...
794.61	5	794.61		...
808.59	10	808.59		...
810.60	25	810.60		...
811.61	20	811.61		...
813.69	5	813.69		...
821.53	5	821.53		...

# Fiehnlab LipidBlast uses NIST MS in batch queries

(Tobias Kind)

NIST MS Search 2.0 - [Peptide, Presearch Default - 42 spectra]

File Search View Tools Options Window Help

1. 732.555 [Da].dta

**Experimental MS/MS list**

#	Src.	Name
32	A	758.571 [Da].dta
33	A	759.573 [Da].dta
34	A	760.586 [Da].dta
35	A	762.599 [Da].dta
36	A	766.536 [Da].dta
37	A	768.555 [Da].dta

pc-pos-h; 5476 total spectra

**Library hit scores**

#	Li...	Score	Dot Pro...	Prob. ...	E-Om...	Name
1	pc	855	855	25.0	0	PC 32:1; [M+H] <sup>+</sup> ; GPCho(1)
2	pc	855	855	25.0	0	PC 32:1; [M+H] <sup>+</sup> ; GPCho(1)
3	pc	855	855	25.0	0	PC 32:1; [M+H] <sup>+</sup> ; GPCho(1)
4	pc	855	855	25.0	0	PC 32:1; [M+H] <sup>+</sup> ; GPCho(1)
5	pc	106	106	0.00	0	PC 32:1; [M+H] <sup>+</sup> ; GPCho(2)
6	pc	106	106	0.00	0	PC 32:1; [M+H] <sup>+</sup> ; GPCho(8)
7	pc	75	75	0.00	0	PC 32:1; [M+H] <sup>+</sup> ; GPCho(1)
8	pc	75	75	0.00	0	PC 32:1; [M+H] <sup>+</sup> ; GPCho(1)
9	pc	69	69	0.00	0	PC 32:1; [M+H] <sup>+</sup> ; GPCho(2)
10	pc	69	69	0.00	0	PC 32:1; [M+H] <sup>+</sup> ; GPCho(6)
11	pc	61	61	0.00	0	PC 32:1; [M+H] <sup>+</sup> ; GPCho(1)
12	pc	61	61	0.00	0	PC 32:1; [M+H] <sup>+</sup> ; GPCho(1)
13	pc	54	54	0.00	0	PC 32:1; [M+H] <sup>+</sup> ; GPCho(2)
14	pc	54	54	0.00	0	PC 32:1; [M+H] <sup>+</sup> ; GPCho(1)
15	pc	49	49	0.00	0	PC 32:1; [M+H] <sup>+</sup> ; GPCho(2)
16	pc	49	49	0.00	0	PC 32:1; [M+H] <sup>+</sup> ; GPCho(2)

**exp. MS/MS**

Name: 732.555 [Da].dta  
MW: N/A ID#: 7803 DB: Text File  
Comment: CHARGE=1+ PEPMASS=732.554687  
10 largest peaks:  
673.3751991201706 999.00 | 672.382010  
478.37586510016934 185.76 | 714.5033631  
263 m/z Values and Intensities:  
213.14963098738727 8.26 | 270.37082386  
319.19541630653885 6.65 | 337.2953131  
367.4354270383564 0.59 | 364.1086965

(Text File) 732.555 [Da].dta

**exp. MS/MS**

**in-silico MS/MS**

732.555 [Da].dta Head to Tail MF=438 RMF=901 PC 32:1; [M+H]<sup>+</sup>; GPCho(16:1(9)).

**in-silico MS/MS**

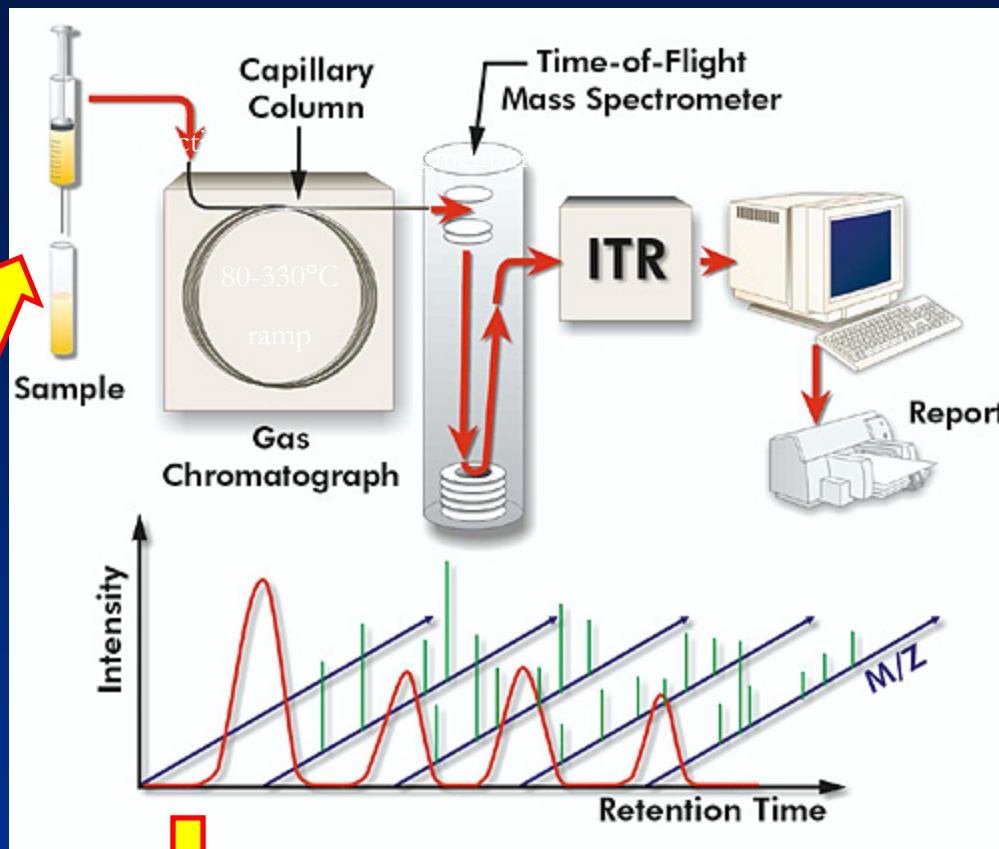
Name: PC 32:1; [M+H]<sup>+</sup>; GPCho(16:1(9Z)/16:0)  
MW: 732 ID#: 3924 DB: pc-pos-h  
7 m/z Values and Intensities:  
476.31425 600.00 [M+H]-sn2-H2O  
478.32989 600.00 [M+H]-sn1-H2O  
494.32481 600.00 [M+H]-sn2  
496.34045 600.00 [M+H]-sn1  
549.48829 400.00 [M+H]-C5H14NO4P (-183)  
673.48083 999.00 [M+H]-C3H9N (-59)  
714.54377 400.00 [M+H]-H2O (-18)

For Help, press F1

Peptide

# Metabolomics for volatiles (Kirsten Skogerson)

('aroma, fragrances, biofuels, lung diseases...')



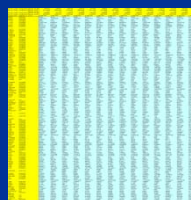
Trap volatiles



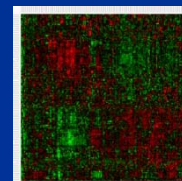
Desorb volatiles onto GC column



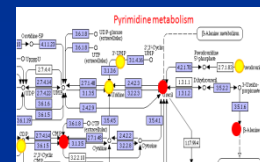
Fiehnlab BinBase DB



Statistics

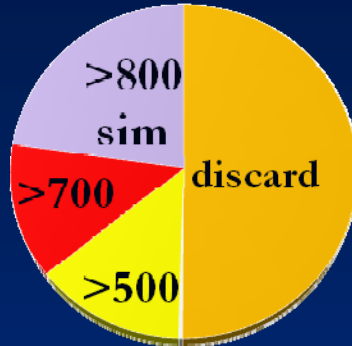
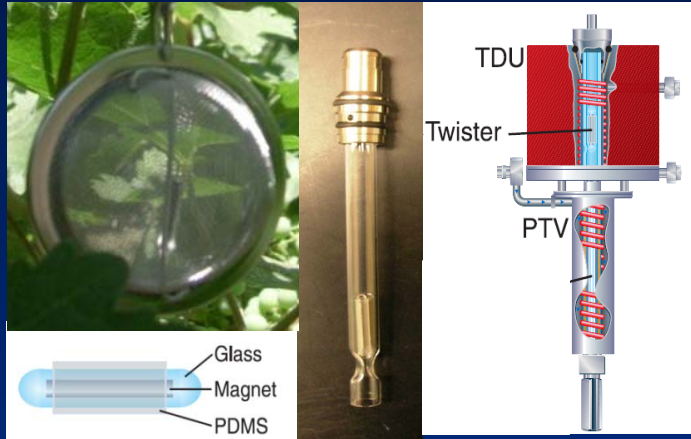


Mapping

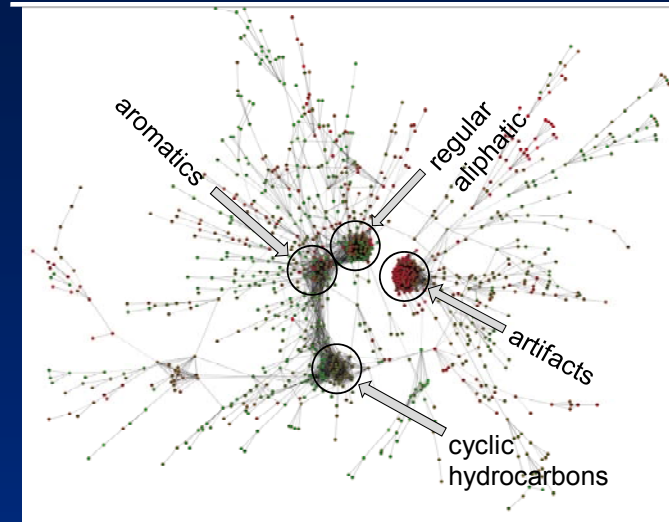




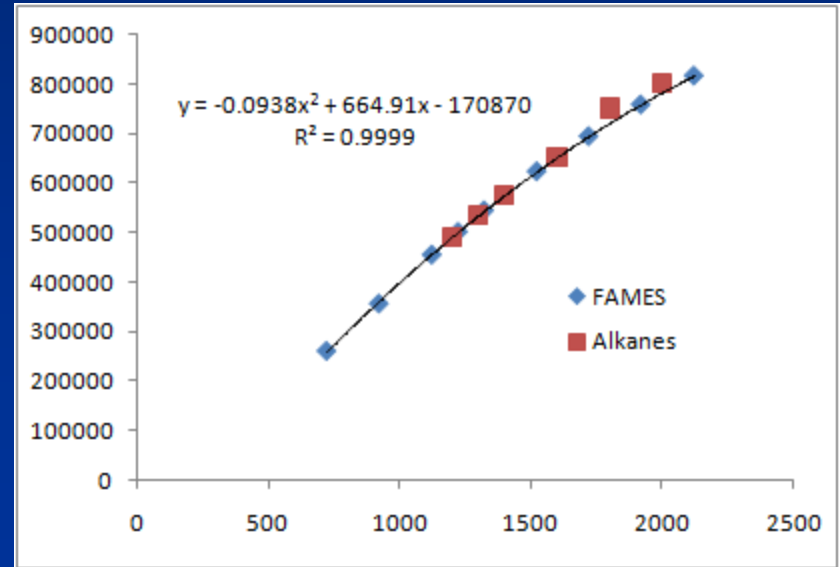
# Twister-TDU-GC-TOF for volatile profiling



BinBase annotations  
 1,500 samples  
 1,000 unique volatiles

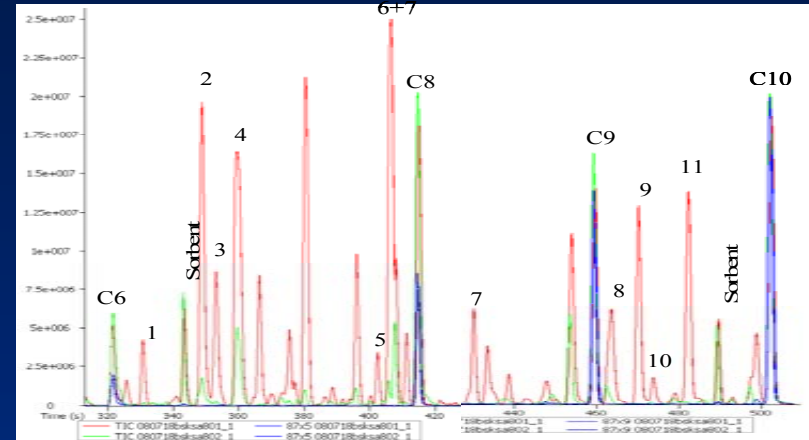
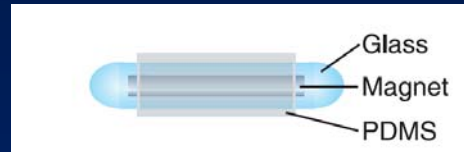


Compound	Adams Ret.Index	Fiehn Ret.Index
1.63 Ethyl ether	529	154618
1.80 Ethyl acetate	606	197619
1.88 Hexane<n->	623	206962
1.93 Methyl propanoate	634	212979
1.99 Dichloroethane<1,2->	647	220061
2.00 Cyclobutanone	650	221691
2.00 Isobutanol	650	221691
2.02 Butanal<2-methyl->	654	223861
2.04 Isovaleraldehyde	658	226029
2.10 Methyl-2-butanol<3->	671	233052
2.14 Isobutyl formate	680	237896
2.15 Pentanone<2->	682	238970
2.15 Pyruvic acid	682	238970
2.16 Penten-3-ol<1->	684	240044
2.18 Pentanol<2->	689	242724
2.23 Heptane	700	248605
2.23 Pentandione<2,3->	700	248605
2.25 Ethyl furan<2->	702	249672



# Volatile BinBase

1,200 volatile Twister-CIS-GCTOF profiles for in-field monitoring of grape maturity, Napa Valley

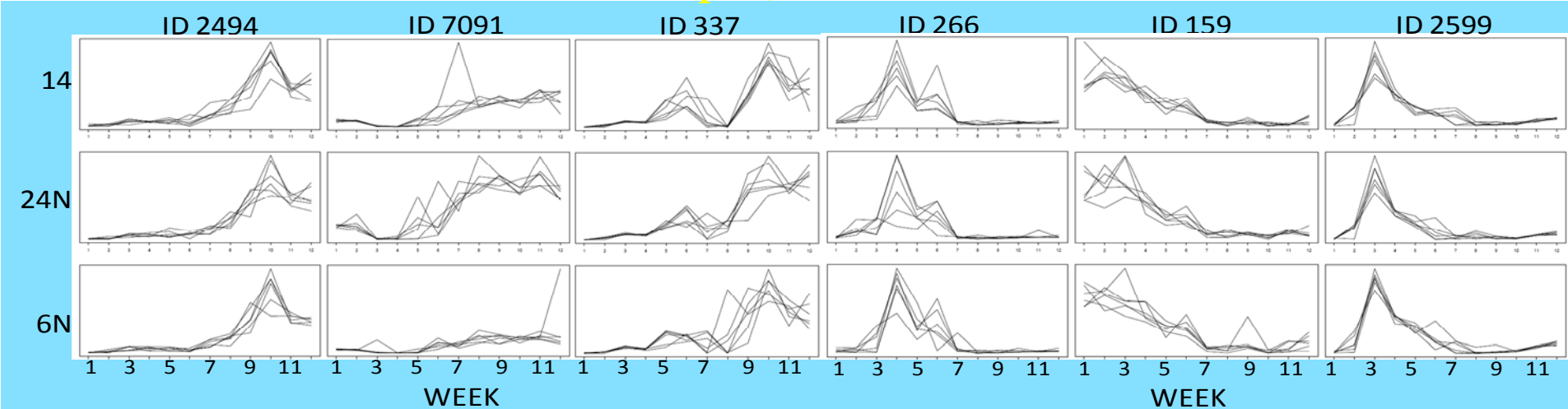


1.  $\beta$ -pinene, 2. sabinene, 3.  $\alpha$ -pinene /  $\alpha$ -myrcene,
4. 3-hexen-1-ol-acetate, 5. limonene, 6. linalool / nonanal,
7. citronellal, 8. nerol, 9. geraniol, 10. neral, 11. geranial

Adams MS-RI library, 2000 volatiles 5%Ph

Twister

BinBase output, stats

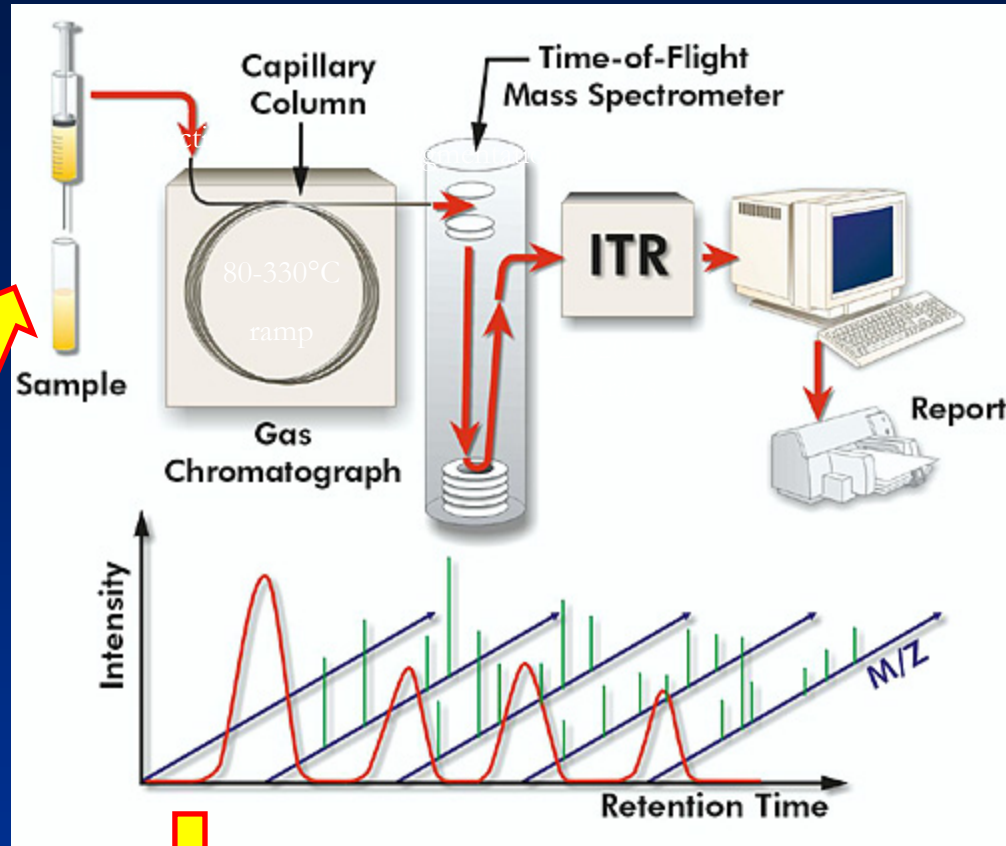


# Metabolomics for primary metabolism

20 mg tissue with homogenization ,  
10 µl plasma, urine

Cold extraction  
(iPrOH, ACN, water)

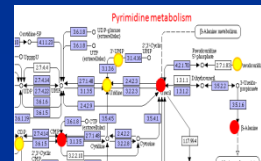
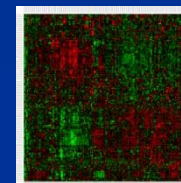
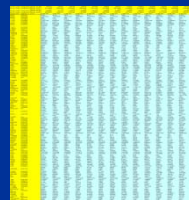
Dry down, derivatize  
to increase volatility



Fiehnlab BinBase DB

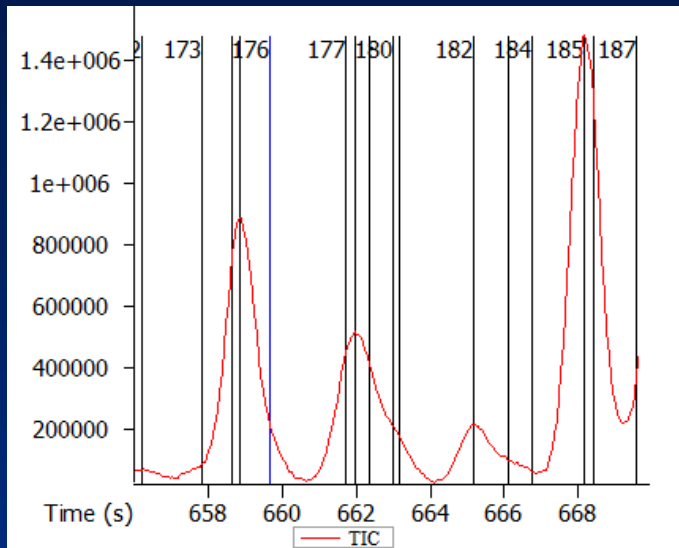
Statistics

Mapping

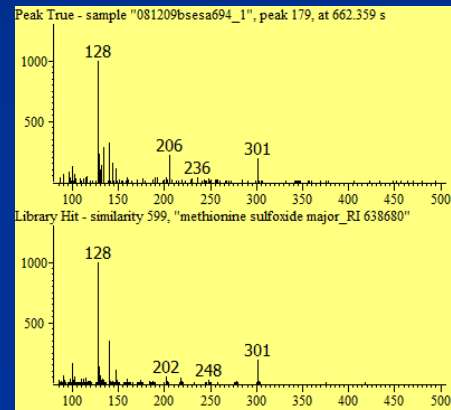
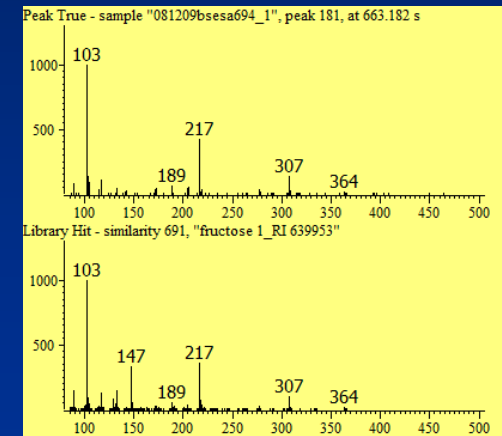
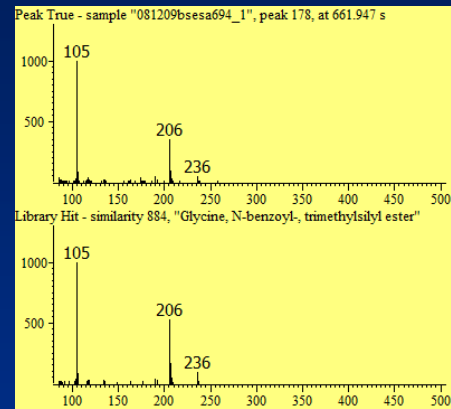
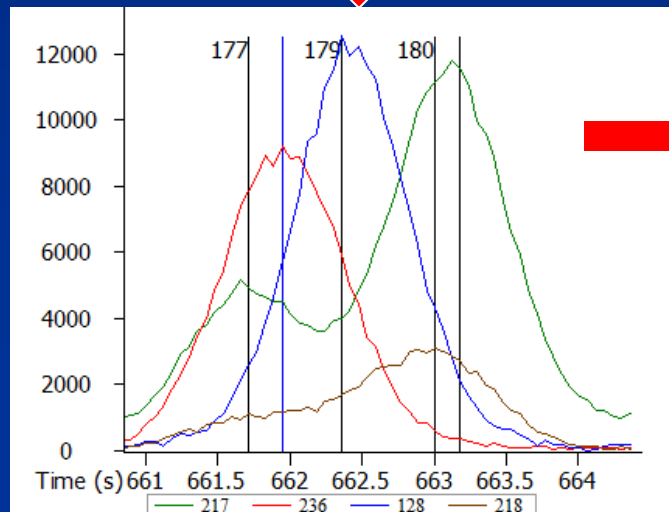


# Peak picking and mass spectral deconvolution is critical for metabolomics

lack of good software for LC-MS  
GC-MS has better software

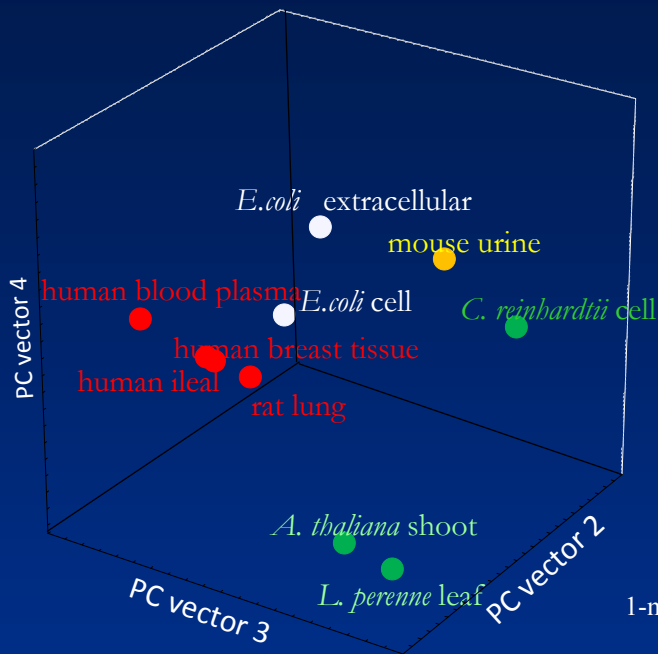


5 peaks in 2 s



# BinBase query of 846 GC-TOF samples in 10 studies

(G. Wohlgemuth)



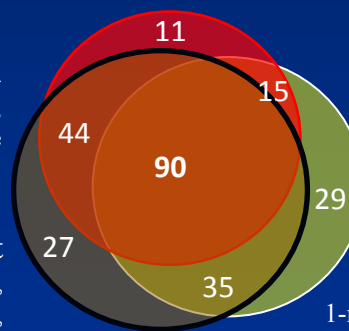
Multivariate PCA on median levels of 338 identified metabolites.

**plasma, urine & ileal effluent**  
e.g. inositol, alanine, palmitate, uric acid, creatinine, glucose, glycerol-alpha-phosphate

**blood plasma**  
e.g. kynurenine, cysteine-glycine, glycerol-beta-phosphate, azelaic acid

**plasma & ileal**  
e.g. cholesterol, gamma-tocopherol, 1-monopalmitin, glycerol, arachidonate

**plasma & urine**  
e.g. conduritol-b-expoxide, mannitol, 2-deoxyerythritol, 2-phenylpropanol

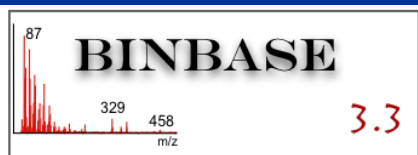


**ileal effluent**  
e.g. beta-sitosterol, cholic acid, spermidine, nonadecanoic acid, guanosine

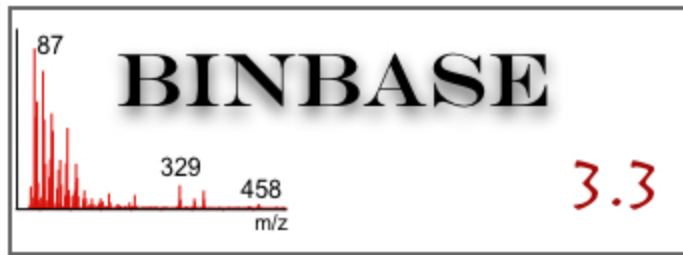
**urine**  
e.g. 2-ketoadipate, citramalate, 1-methyladenosine, phenylpyruvate, N-hexanoylglycine

**ileal & urine**  
e.g. 5-aminovalerate, glutarate, uracil, ribose, glycerol-3-galactoside

Venn-diagram of qualitative overlap of identified metabolites in the three mammalian biofluids.



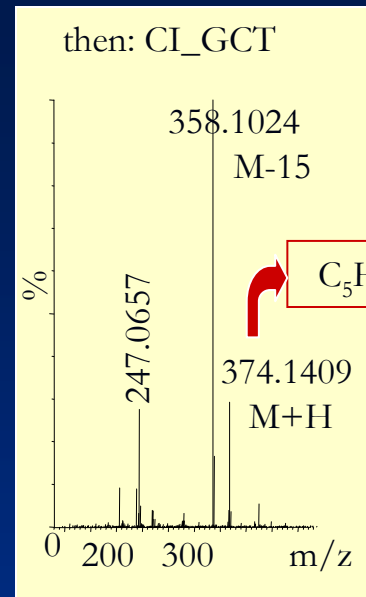
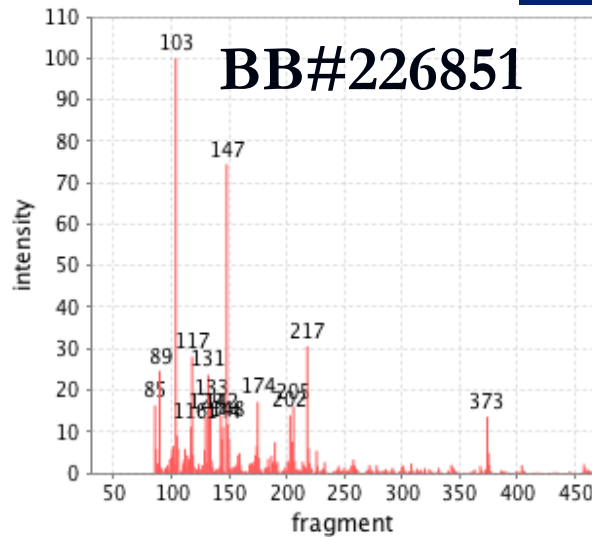
# Public query for BinBase metabolites by spectra or ID



Change Database | Search Bins | List Species | BinBase Publications

Search for Bin's by Name

Similarity Search for Bin



PubChem  
ChemSpider  
CSLS DB

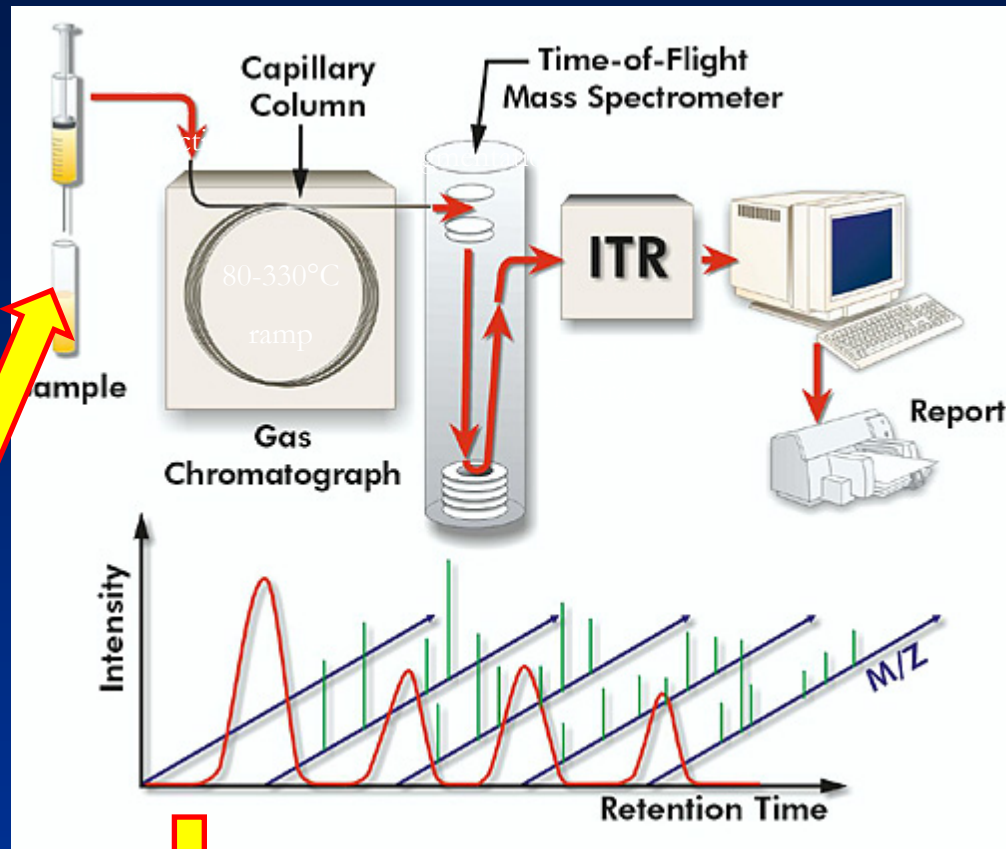
samples species	#199773	#226851	#225396	#22858
1701 Homo sapiens	28%	11%	28%	3
922 Rattus norvegicus	16%	21%	15%	1
922 C. reinhardtii	29%	0%	0%	1
917 Mus sp.	19%	6%	3%	1
88 Ovis aries	94%	99%	15%	9
144 Acyrthosiphon pisum	0%	97%	0%	
434 Saccharophagus degradans	0%	9%	3%	
109 Prunus persica	0%	1%	0%	5
88 E. coli	0%	0%	0%	3
2738 Arabidopsis thaliana	0%	0%	0%	
1408 Solanum tuberosum	0%	0%	0%	
800 Zea mays	0%	0%	0%	
487 Glycine max	0%	0%	0%	
483 Malus domestica	0%	0%	0%	
387 Pinus taeda	0%	0%	0%	
289 Citrus sinensis	0%	0%	0%	
168 Beta vulgaris	0%	0%	0%	
135 Pyrus communis	0%	0%	0%	
129 Lolium perenne	0%	0%	0%	

<http://eros.fiehnlab.ucdavis.edu:8080/binbase-compound/>

Open source, documented.  
105 users, 45 laboratories

# Metabolomics for polymers (Dinesh K Barupal)

(lignins, cellulose, hemicellulose, complex lipids...)



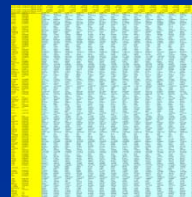
2 mg material



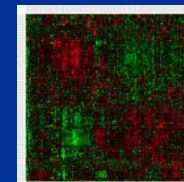
pyrolysis



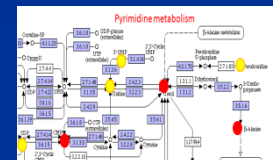
SpectConnect (MIT)



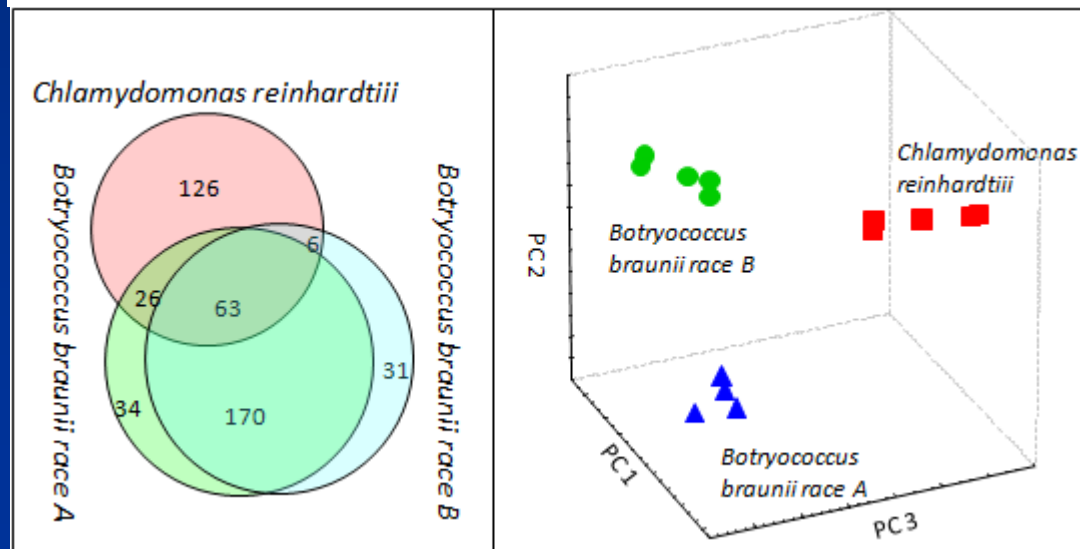
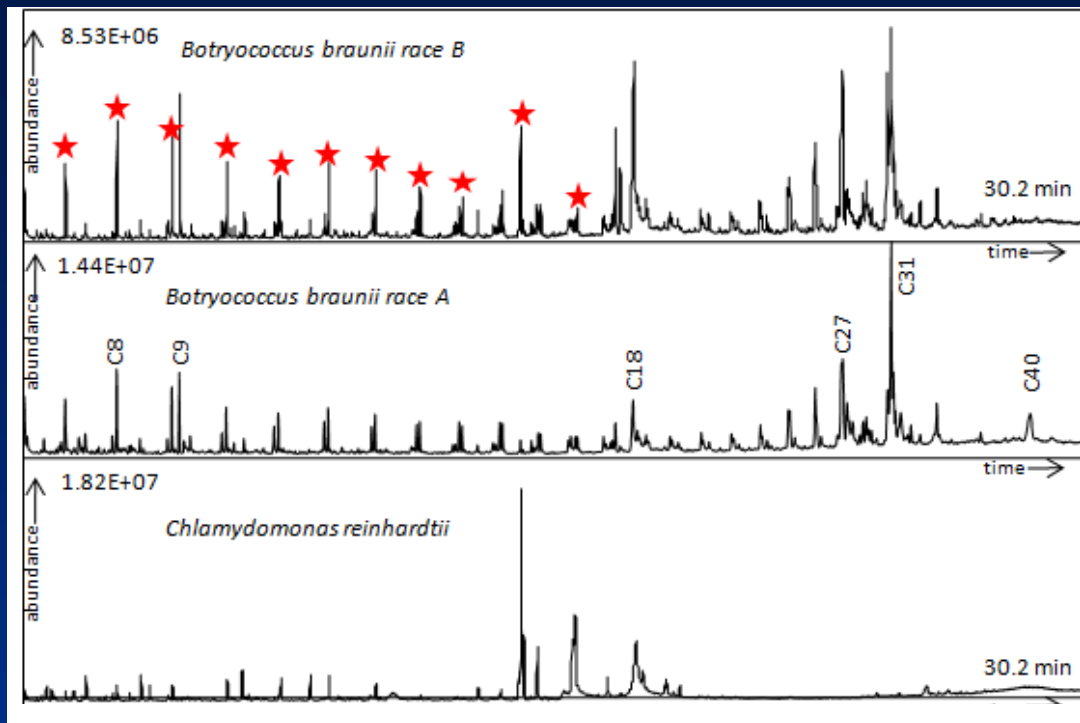
Statistics



Mapping



# pyGC-MS for total lipids in biofuel algae





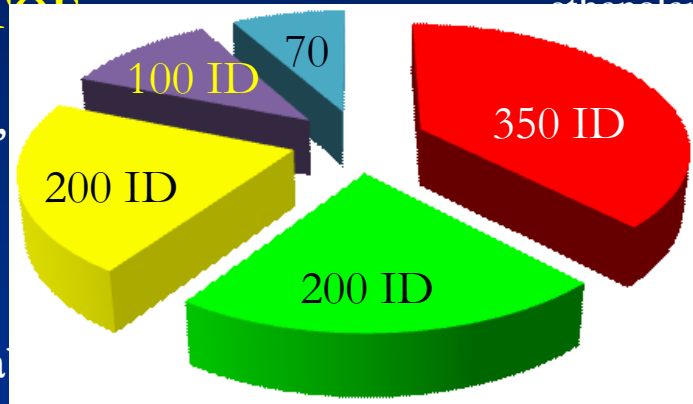


# What about UPLC-MS/MS?

**pyGC-MS**  
monomers  
lignin, hemicellulose  
complex lipids

**nanoESI-MS/MS** polar & neutral lipids  
**UPLC-MS/MS**  
phosphatidylcholines, -serines,  
sphingomyelins, plasmalogens, triglycerides

**Twister-GC-TOF**  
volatiles  
terpenes, alkanes,  
FFA, benzenes



**GCxGC-TOF**  
primary small meta

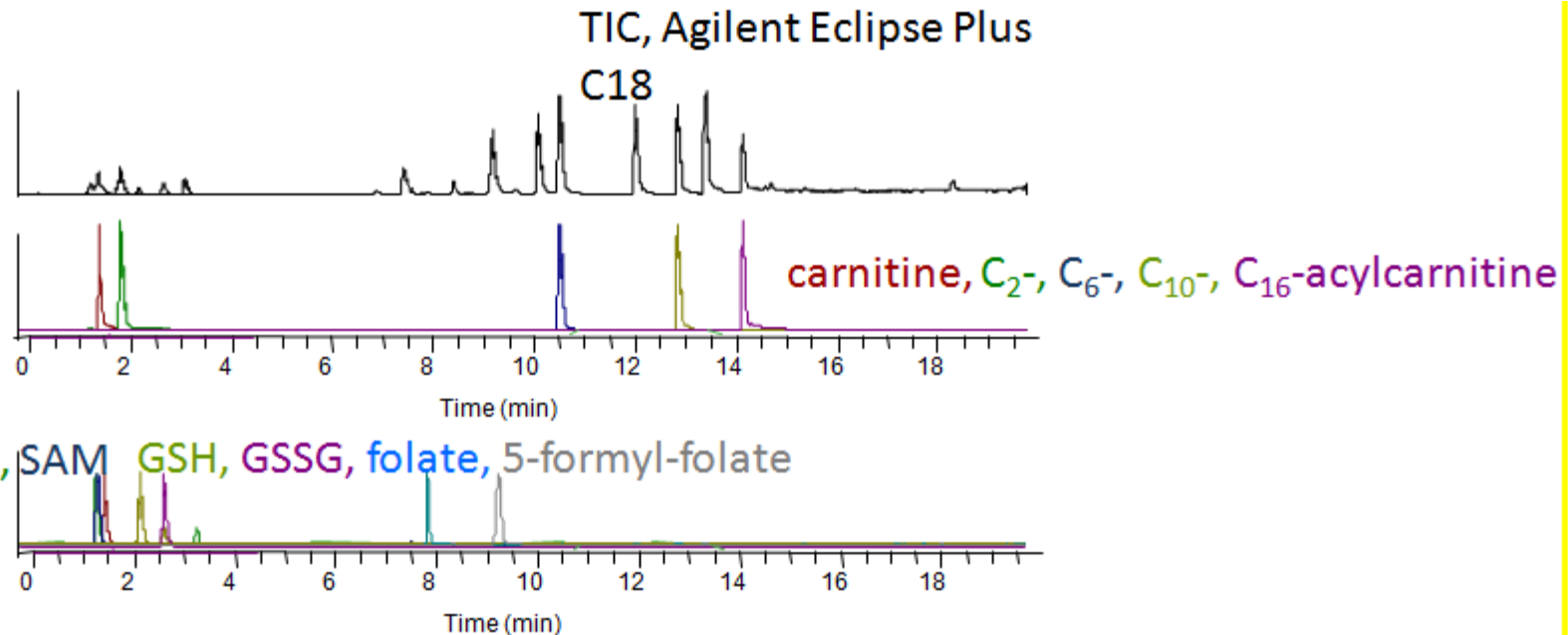
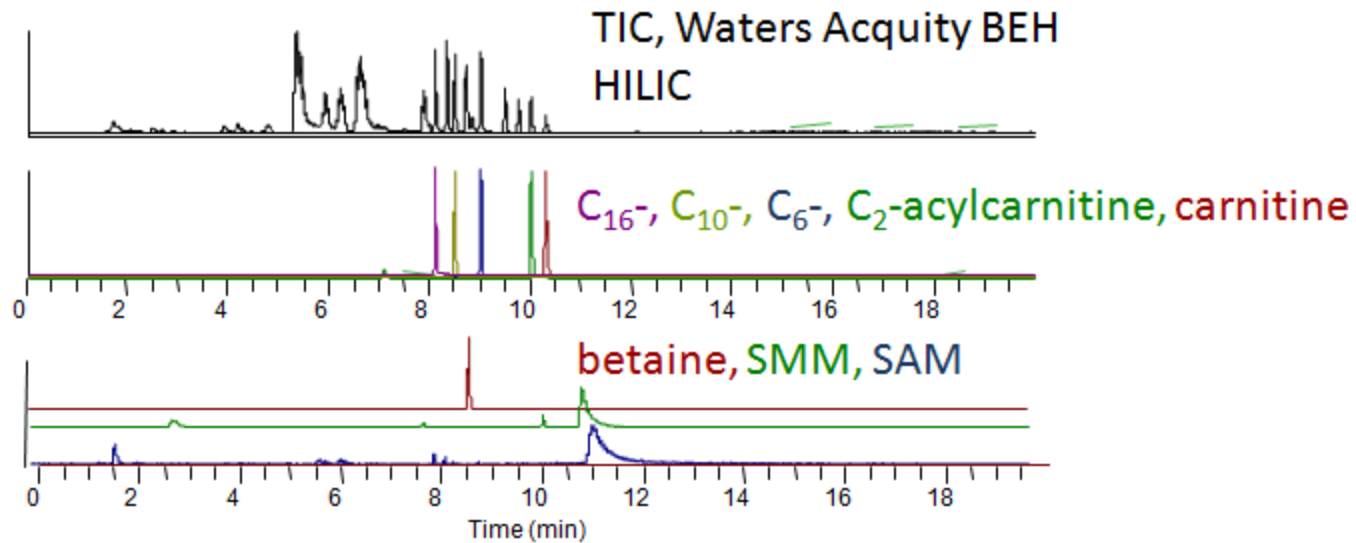
sugars, HO-acids, FFA, amino acids,  
sterols, phosphates, aromatics

**UPLC-UV-MS/MS** secondary metabolites

oxylipids, anthocyanins, flavonoids, pigments  
acylcarnitines, folates, glucuronidated & glycosylated aglycones

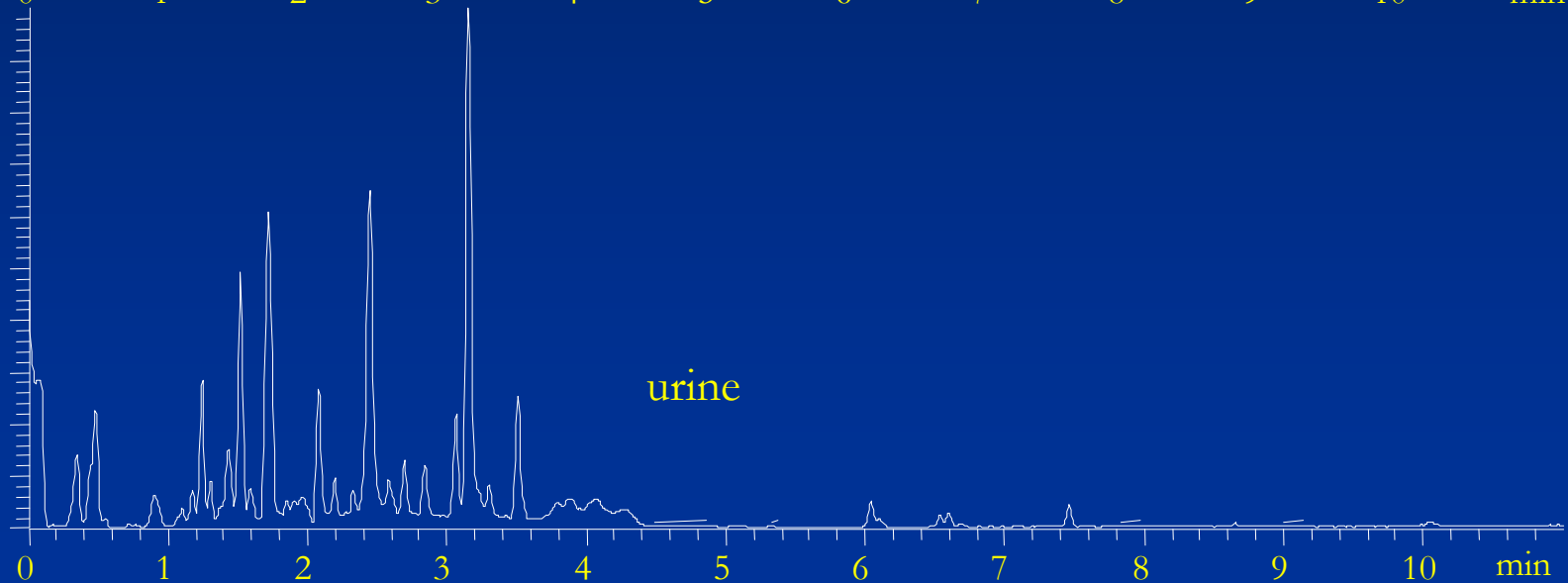
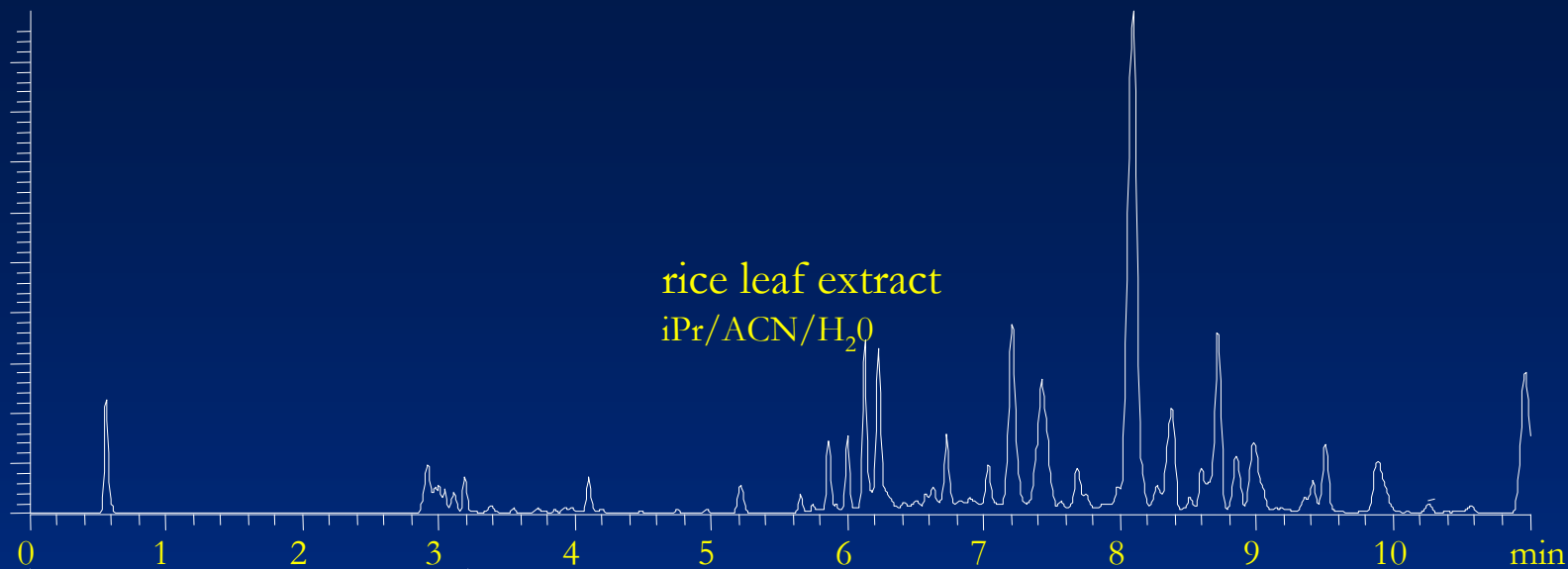
# Which LC-MS/MS method?

(John Meissen)



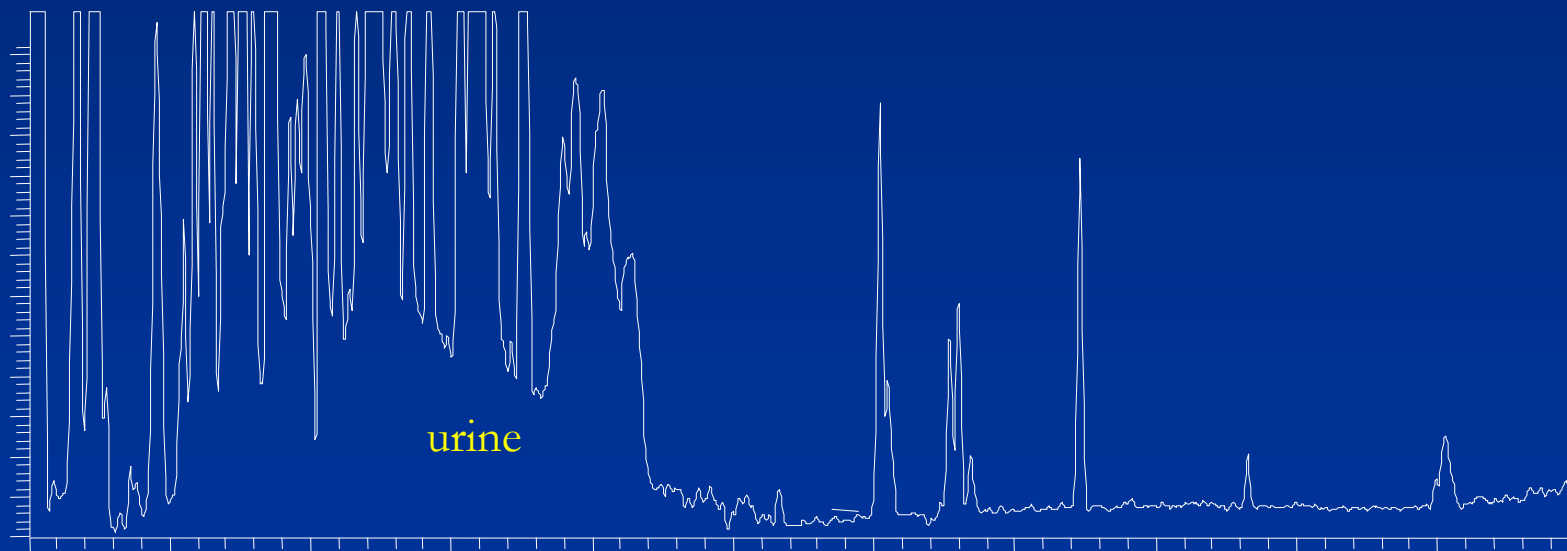
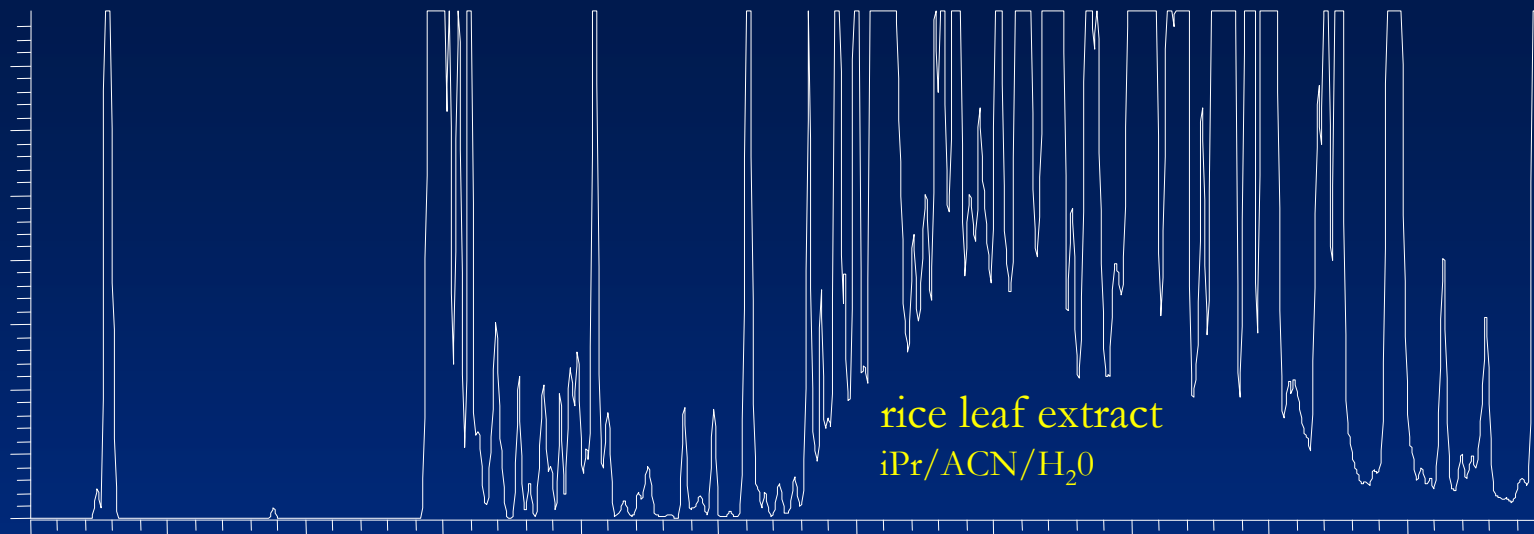
# Which LC-MS/MS method?

V. Tolstikov, UC Davis Metabolomics Core



C18 1.7 $\mu$ m particles column 2.1x100 mm at 40°C , 0.5 ml/min NH<sub>4</sub>OAc 13 mM pH 5.5 to 100% ACN. NegESI 100-1500 Da

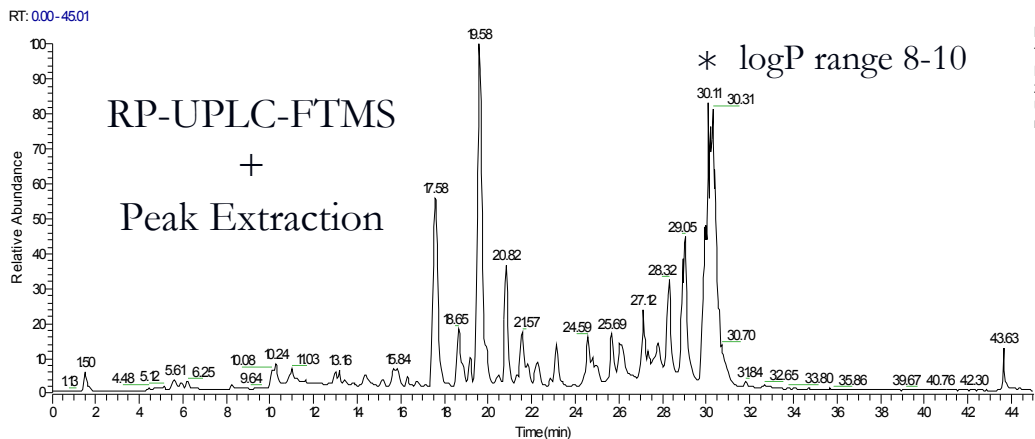
# Lack of peak capacity



Even if there was good peak picking, MS deconvolution, accurate masses:  
Can we get correct elemental compositions?

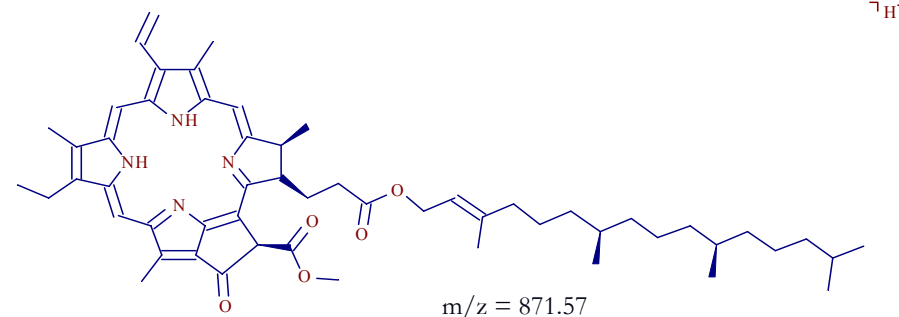
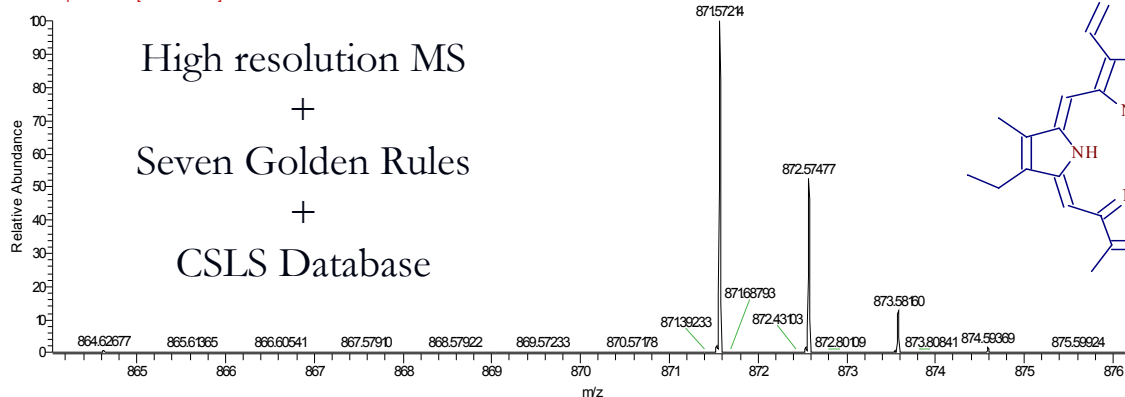
## 7 Golden Rules (Tobias Kind)



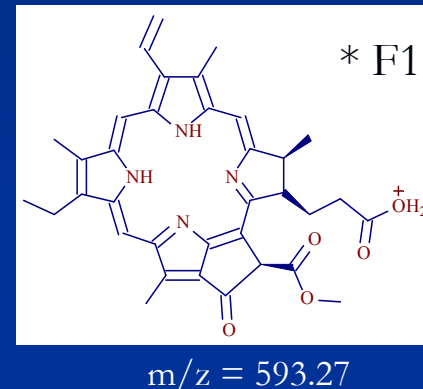
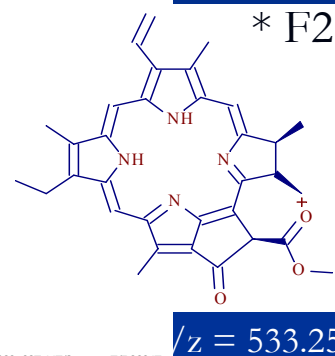
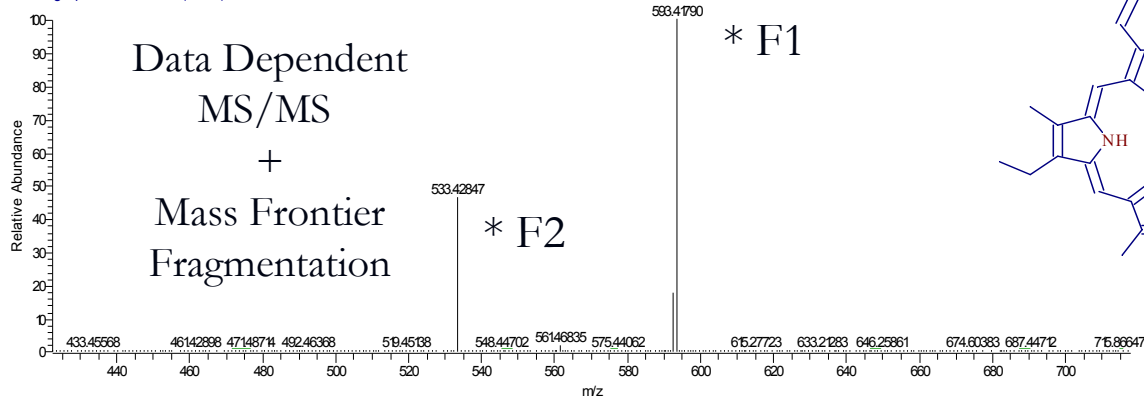


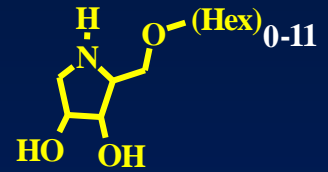
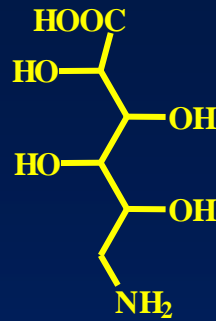
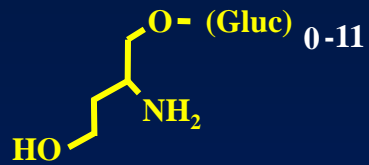
# MS/MS confirmation by matching MassFrontier predicted spectra

UPLC\_C8\_DataDependent\_Rical309mock #2935 RT: 30.05 AV: 1 NL: 282E6  
F: FTMS+pNSI Full ms [200.00-1200.00]

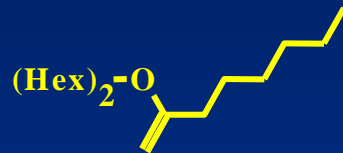


UPLC\_C8\_DataDependent\_Rical309mock #8-4240 RT: 0.114388 AV: 81 NL: 623E4  
T: Average spectrum MS287157 (8-4240)

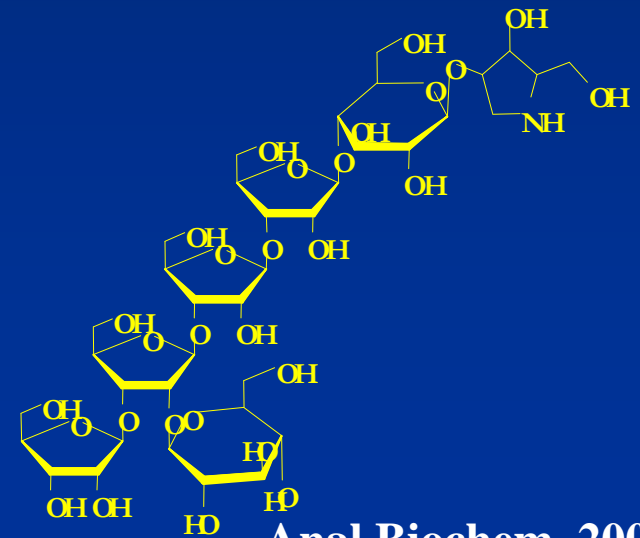
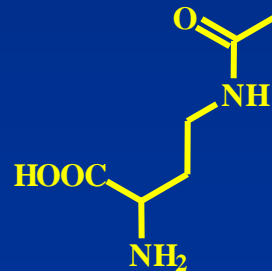
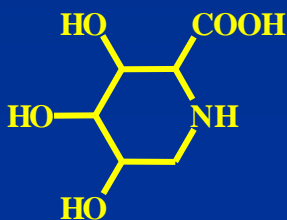
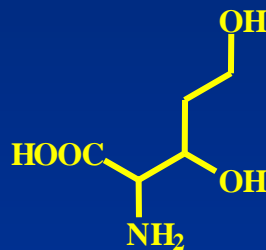
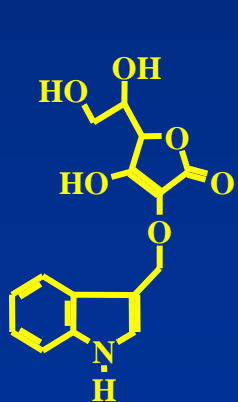




„In reality, the majority of detected components are unidentified compounds with unknown functions that are part of undetermined biochemical pathways.“



Vladimir Tolstikov



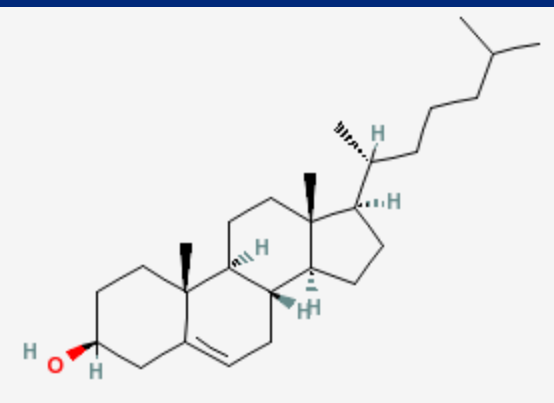


# There a unique way to name compounds.

CAS numbers are **non-public**, not unique identifiers.

Unambiguous is the **chemical structure**

Second best is database identifiers (**PubChem CID**)



## *Open source InChI code and hash keys*

**InChI**=1S/C27H46O/c1-18(2)7-6-8-19(3)23-11-12-24-22-10-9-20-17-21(28)13-15-26(20,4)25(22)14-16-27(23,24)5/h9,18-19,21-25,28H,6-8,10-17H2,1-5H3/t19-,21+,22+,23-,24+,25+,26+,27-/m1/s1

**InChIKey**: HVYWMOMLDIMFJA-DPAQBDIFSA-N

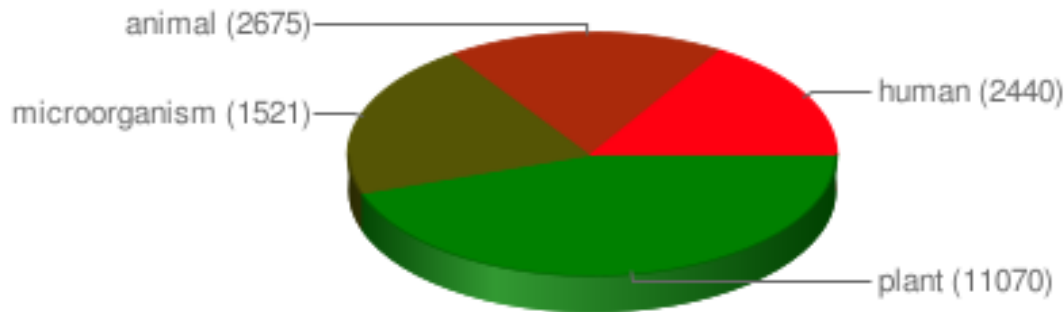


>22,200 samples (Feb 2010)

ID	class	label	comment
597967	597988	sx381227	-your comment-
597971	597988	sx381228	-your comment-

add additional fields:  
MSI Attributes

- MSI Attributes-
- MSI Attributes-
- Genotype
- Organ
- Organ specification
- Cell type
- Subcellular location
- BioSource amount
- Growth support
- Growth location
- Growth plot design
- Humidity at day
- Temperature at night
- Watering regime
- Nutritional regime
- Date(s) of plant: transplanting
- Harvest date, time
- Plant growth stage
- Metabolism quenching methods
- Harvest method
- Sample storage



### Public Data

A subset of 1980 out of the 18024 samples related to 261 experiments is publicly available - including the whole experimental design and the annotated GC-TOF result data. Use the link below to query the whole system for the samples that you are interested in.

RSS



Open source, documented

<http://fiehnlab.ucdavis.edu:8080/m1/login.jsp>

Maintained by core services

# Conclusion: Metabolomics is mature for research!

## Five technology platforms:

- (1) primary metabolism
- (2) secondary metabolism
- (3) membrane lipids
- (4) volatiles
- (5) monomers

## Tools @ [fiehnlab.ucdavis.edu](http://fiehnlab.ucdavis.edu)

- (a) Chem Translator: InChI, PubChem CIDs
- (b) FiehnLib libraries for LipidBlast, GCquad, GCTOF
- (c) SetupX for metadata
- (d) BinBase for MS processing
- (e) Mapping by combining biochem & chem. DB queries

# Thanks to [fiehnlab.ucdavis.edu](http://fiehnlab.ucdavis.edu) !

**Gert Wohlgemuth**  
**Pradeep Hildaya**  
**Tobias Kind**  
**Sevini Shahbaz**  
**Mine Palazoglu**  
**Dinesh Kumar Barupal**  
**Kristie Cloos**  
**Kirsten Skogerson**  
**John Meissen**  
**Pierre Ayotte**  
**Sangeeta Kumari**



**UC DAVIS**  
University of California, Davis

Sponsors: NIEHS R01 ES13932, NIGMS, NSF 2010, EU FP7  
Agilent, Leco, Waters, DuPont, Monsanto, CA citrus growers, Napa winery, BMS

**Thanks to collaborators :** Vladimir Tolstikov (Metabolomics Core)  
Weiss, Hartman, Dandekar, Pinkerton, Warden (all UC Davis)