Identification of Metabolites tools and methods

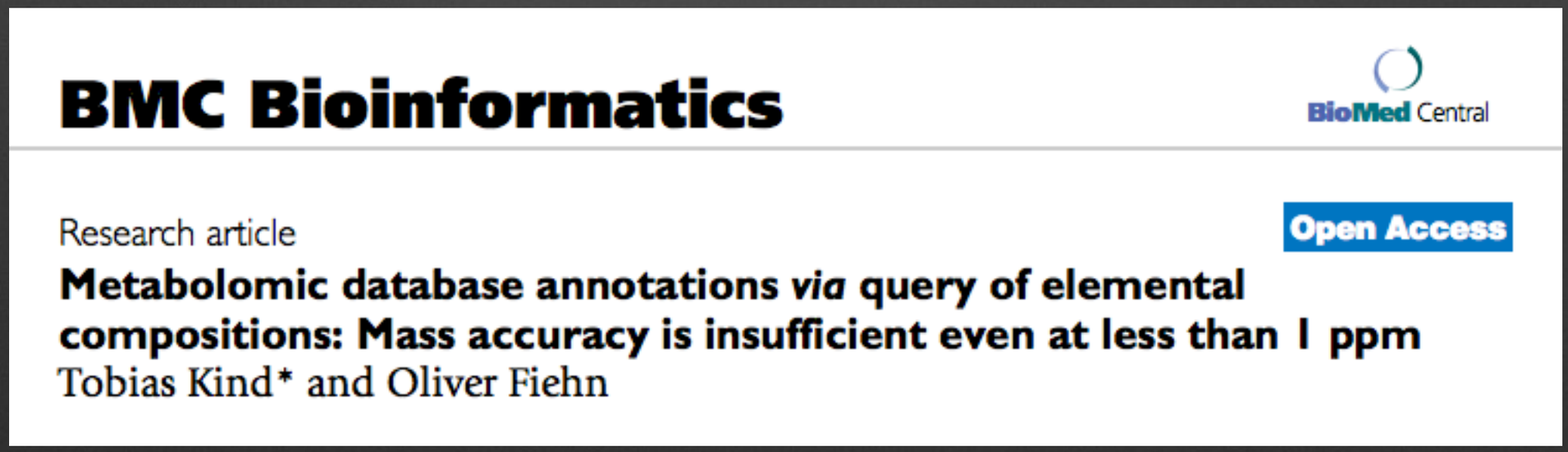
H. Paul Benton

Ways to identify compounds

- Accurate mass - Normally first step to finding the compound - database search using small ppm window
  - Adducts :-{
- Retention time - running a standard or local method database lookup
  - Drift :-{
- Fragmentation patterns -
  - Data - incomplete databases

Accurate mass

- Features must be characterised to known neutral mass
- Mass Spectrometers have errors on mass - leads to larger search space.
- Too many isobaric compounds



BMC Bioinformatics

BioMed Central

Research article

Open Access

Metabolomic database annotations *via* query of elemental compositions: Mass accuracy is insufficient even at less than 1 ppm

Tobias Kind* and Oliver Fiehn

Retention time

- Retention time against standards is a powerful tool
  - Must have the standards
  - Have to keep the same gradient for database lookup method
    - Columns and LC methods are changing
    - Don’t always use LC method - Rapid injection methods

Fragmentation

- Fragmentation is a (sort of) 2D representation of the compound
  - Lots of issues
    - Specificity of isolation window
    - Scoring algorithms
    - Access to standards

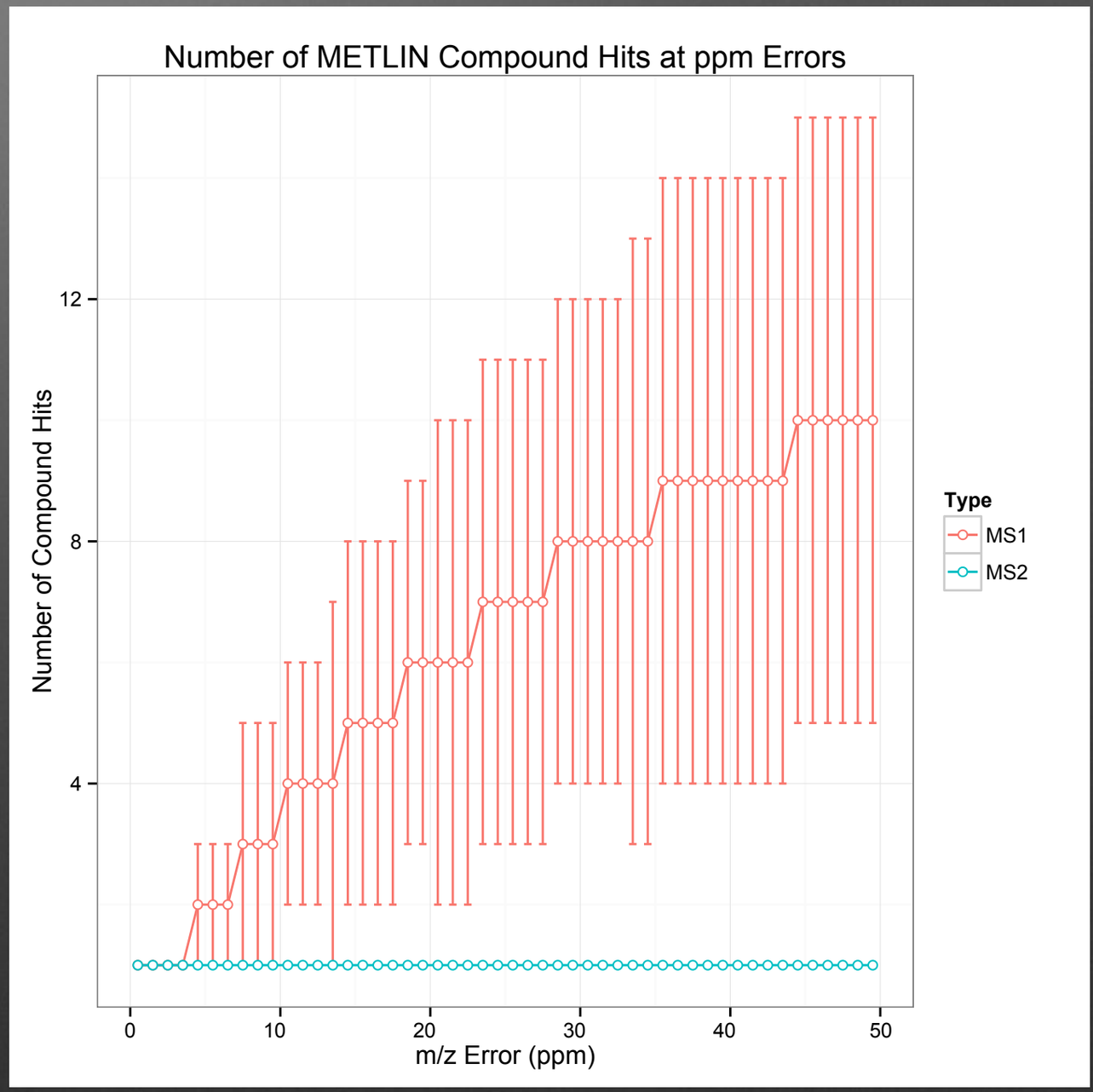
Matching spectra to compounds Scoring methods

- Either matching to a standard or a database
- Many different metrics some are more sensitive
  - Cosine similarity suffers from low resolution issues and should be low sensitivity and specificity
- X-Rank - a very cool algorithm looks at peaks that matter the most
- Many more

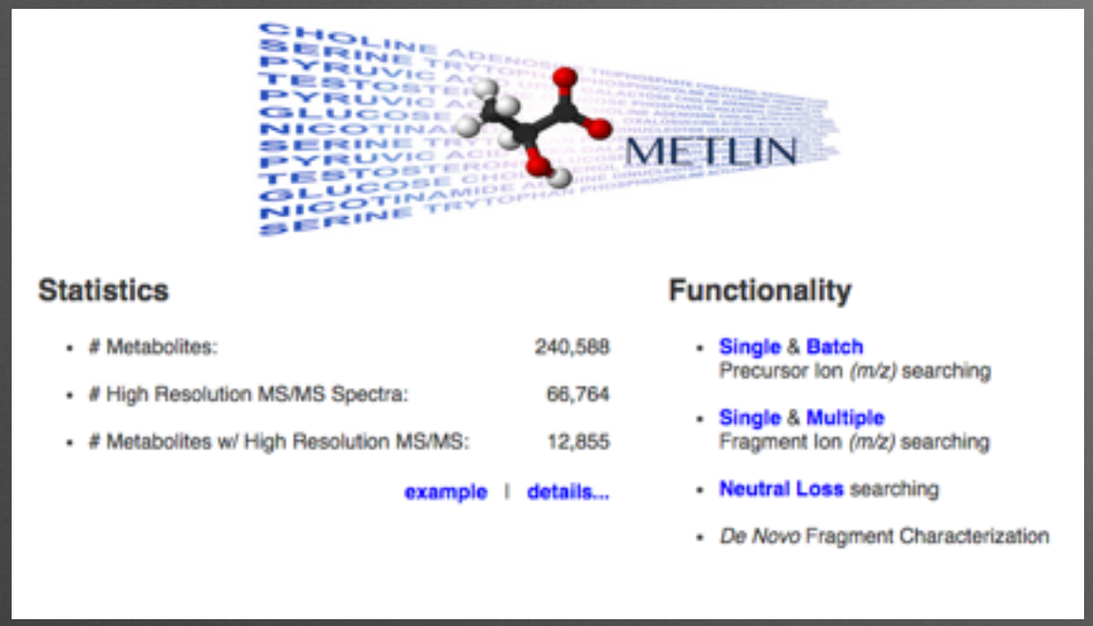
$$\text{similarity} = \cos(\theta) = \frac{A \cdot B}{\|A\|\|B\|} = \frac{\sum_{i=1}^{n} A_i \times B_i}{\sqrt{\sum_{i=1}^{n} (A_i)^2} \times \sqrt{\sum_{i=1}^{n} (B_i)^2}}$$

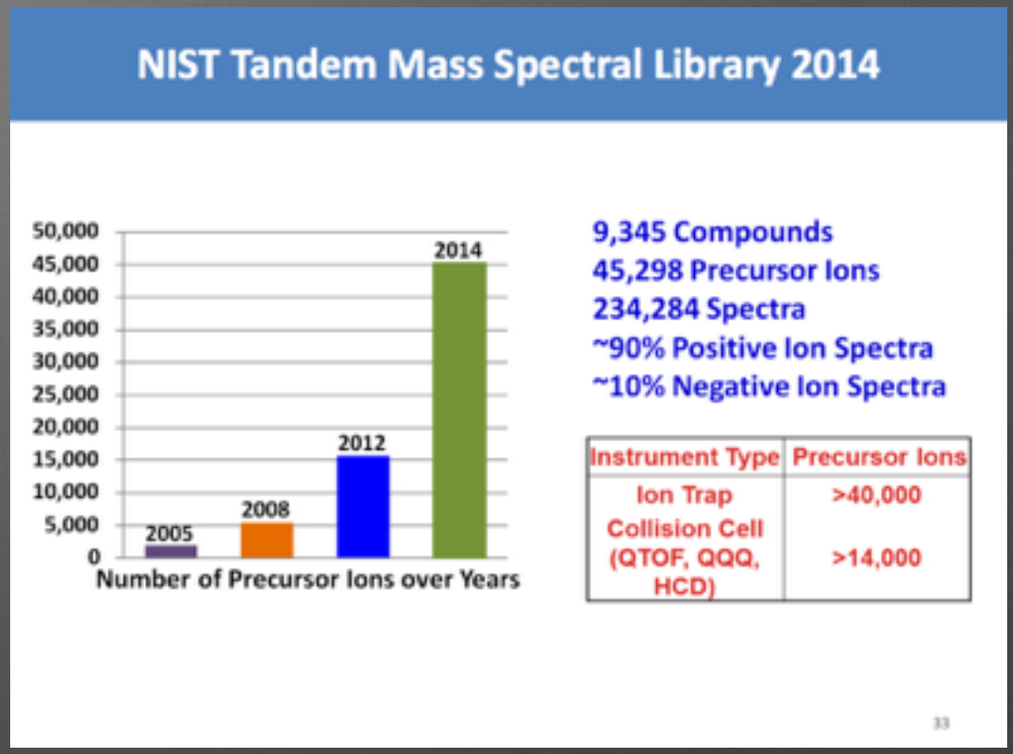
Why it works

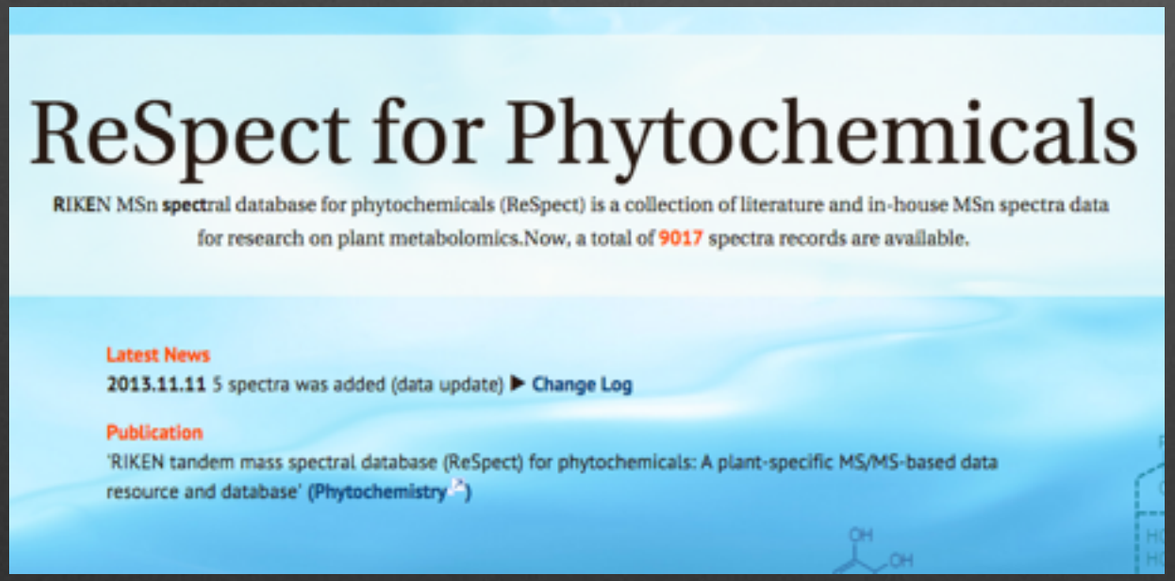
- Precursor selection with fragmentation data is highly specific
- The scoring is only needed for ranking purposes



Databases

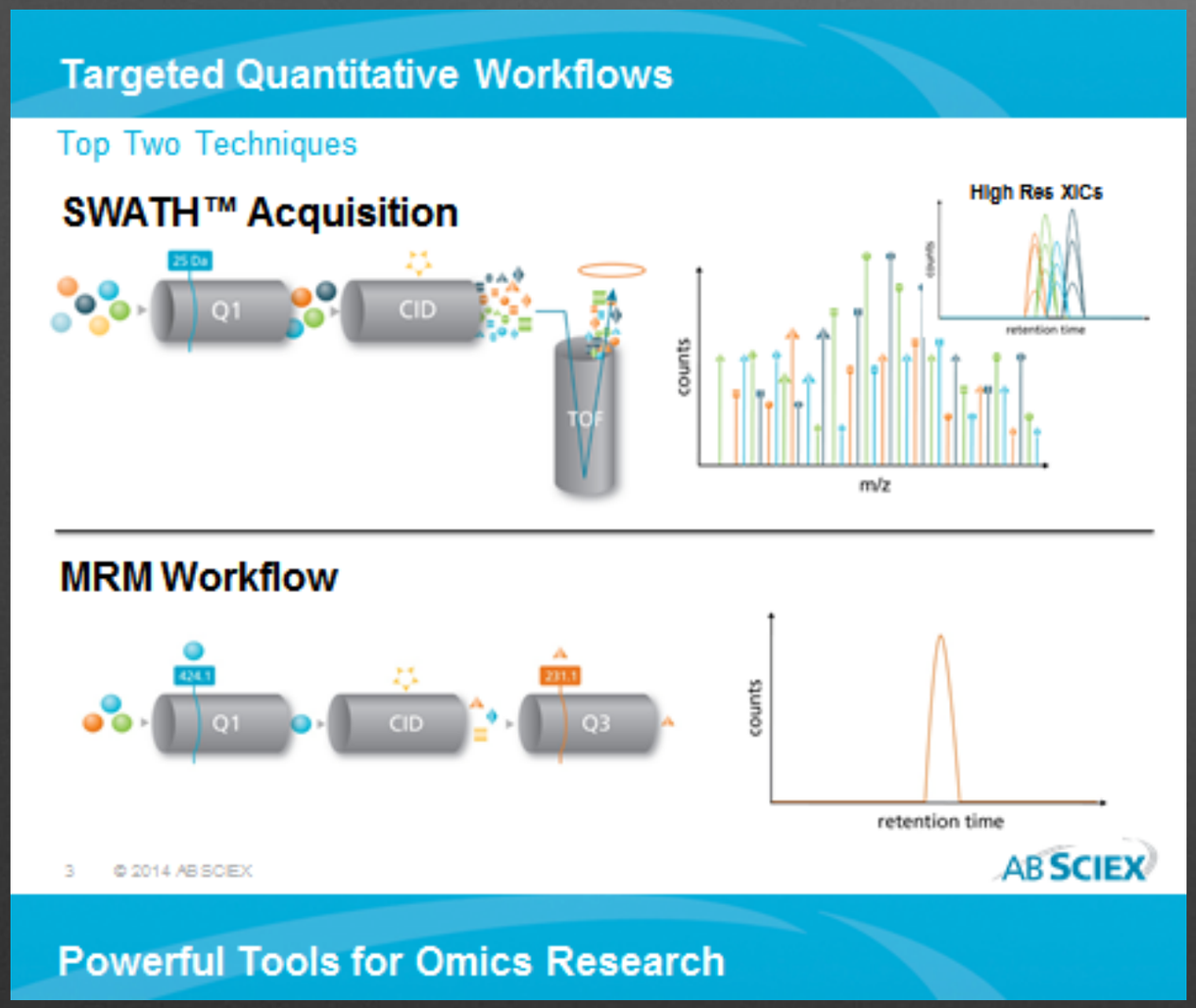






Very hard to tell how many unique compounds Mass Bank has

Methods to matching MS-DIAL

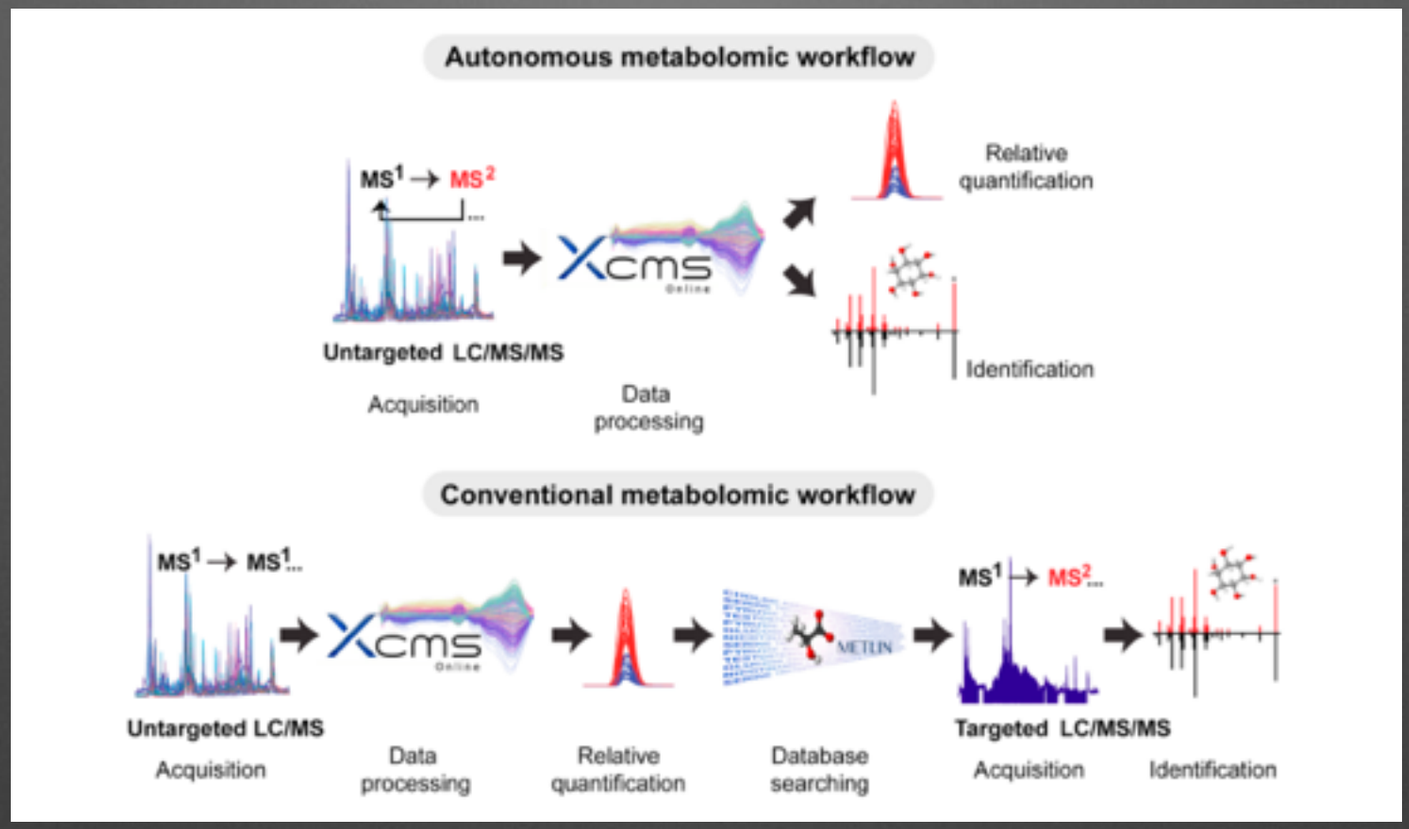


MS-DIAL: data-independent MS/MS deconvolution for comprehensive metabolome analysis

Hiroshi Tsugawa[1,2], Tomas Cajka[3], Tobias Kind[3], Yan Ma[3], Brendan Higgins[4], Kazutaka Ikeda[5,6], Mitsuhiro Kanazawa[7], Jean VanderGheynst[4], Oliver Fiehn[3,8] & Masanori Arita[1,9]

Data-independent acquisition (DIA) in liquid chromatography (LC) coupled to tandem mass spectrometry (MS/MS) provides comprehensive untargeted acquisition of molecular data. We provide an open-source software pipeline, which we call MS-DIAL, for DIA-based identification and quantification of small molecules by mass spectral deconvolution. For a reversed-phase LC-MS/MS analysis of nine algal strains, MS-DIAL using an enriched LipidBlast library identified 1,023 lipid compounds, highlighting the chemotaxonomic relationships between the algal strains.

Methods to matching autonomous



analytical chemistry

Article

pubs.acs.org/ac

Autonomous Metabolomics for Rapid Metabolite Identification in Global Profiling

H. Paul Benton,[†] Julijana Ivanisevic,[†] Nathaniel G. Mahieu,[‡] Michael E. Kurczy,[†] Caroline H. Johnson,[†] Lauren Franco,[§] Duane Rinehart,[†] Elizabeth Valentine,[#] Harsha Gowda,[†,¶] Baljit K. Ubhi,[/] Ralf Tautenhahn,[†,||] Andrew Gieschen,[⊥] Matthew W. Fields,[§] Gary J. Patti,*[,‡] and Gary Siuzdak*

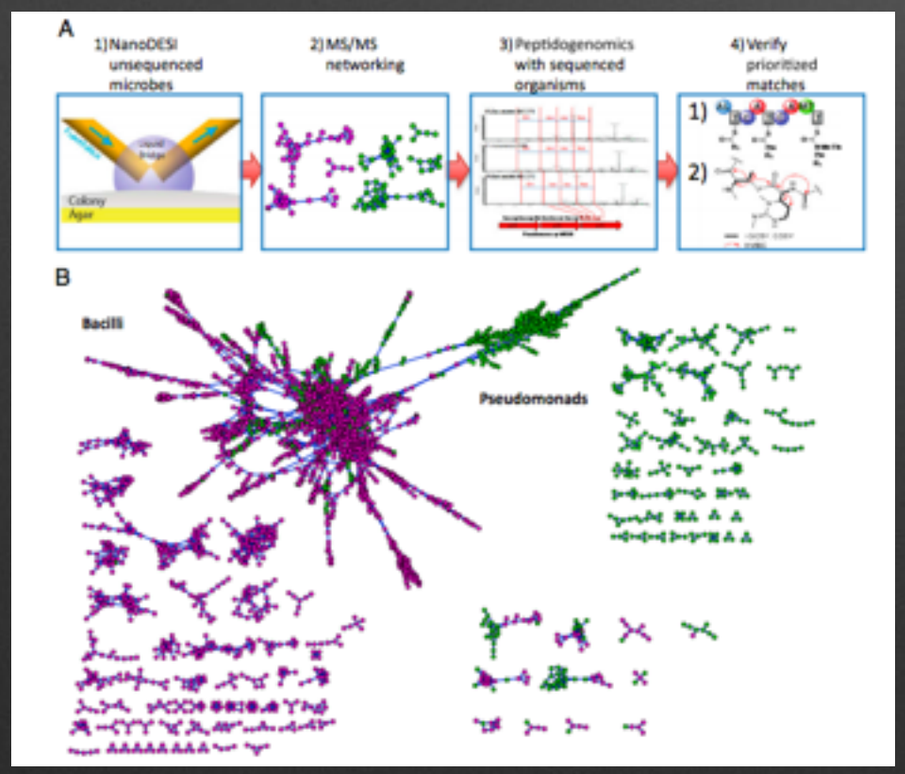
Methods to matching
Similarity matching

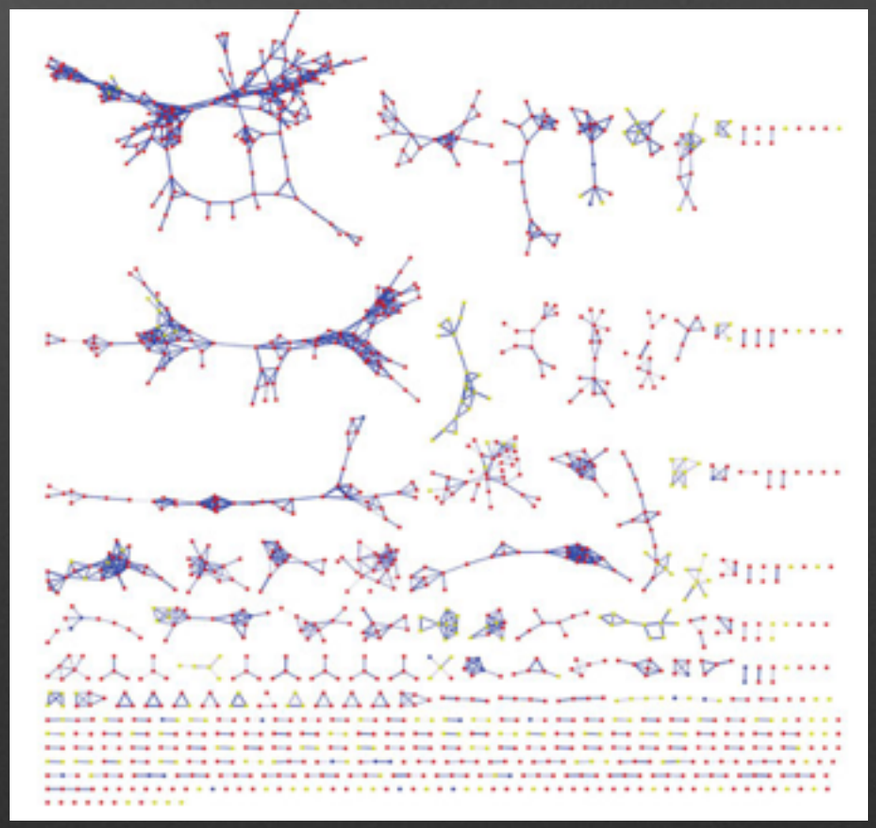
PNAS

MS/MS networking guided analysis of molecule and gene cluster families

Don Duy Nguyen[a,1], Cheng-Hsuan Wu[a,1], Wilna J. Moree[b,1], Anne Lamsa[c], Marnix H. Medema[d], Xiling Zhao[a], Ronnie G. Gavilan[e,f], Marystella Aparicio[e], Librada Atencio[e], Chanaye Jackson[e], Javier Ballesteros[e], Joel Sanchez[e], Jeramie D. Watrous[a], Vanessa V. Phelan[b], Corine van de Wiel[a,b], Roland D. Kersten[g], Samina Mehnaz[h], René De Mot[i], Elizabeth A. Shank[j], Pep Charusanti[k], Harish Nagarajan[k], Brendan M. Duggan[b], Bradley S. Moore[b,g], Nuno Bandeira[b,l,m], Bernhard Ø. Palsson[k], Kit Pogliano[c], Marcelino Gutiérrez[e], and Pieter C. Dorrestein[a,b,g,n,2]

[a]Department of Chemistry and Biochemistry, University of California at San Diego, La Jolla, CA 92093; [b]Skaggs School of Pharmacy and Pharmaceutical





De-novo methods

- General method is to take a structure of a molecule and break bonds (either randomly or rule based) to match to fragments masses.

MS Fragmentor

ACD/Labs

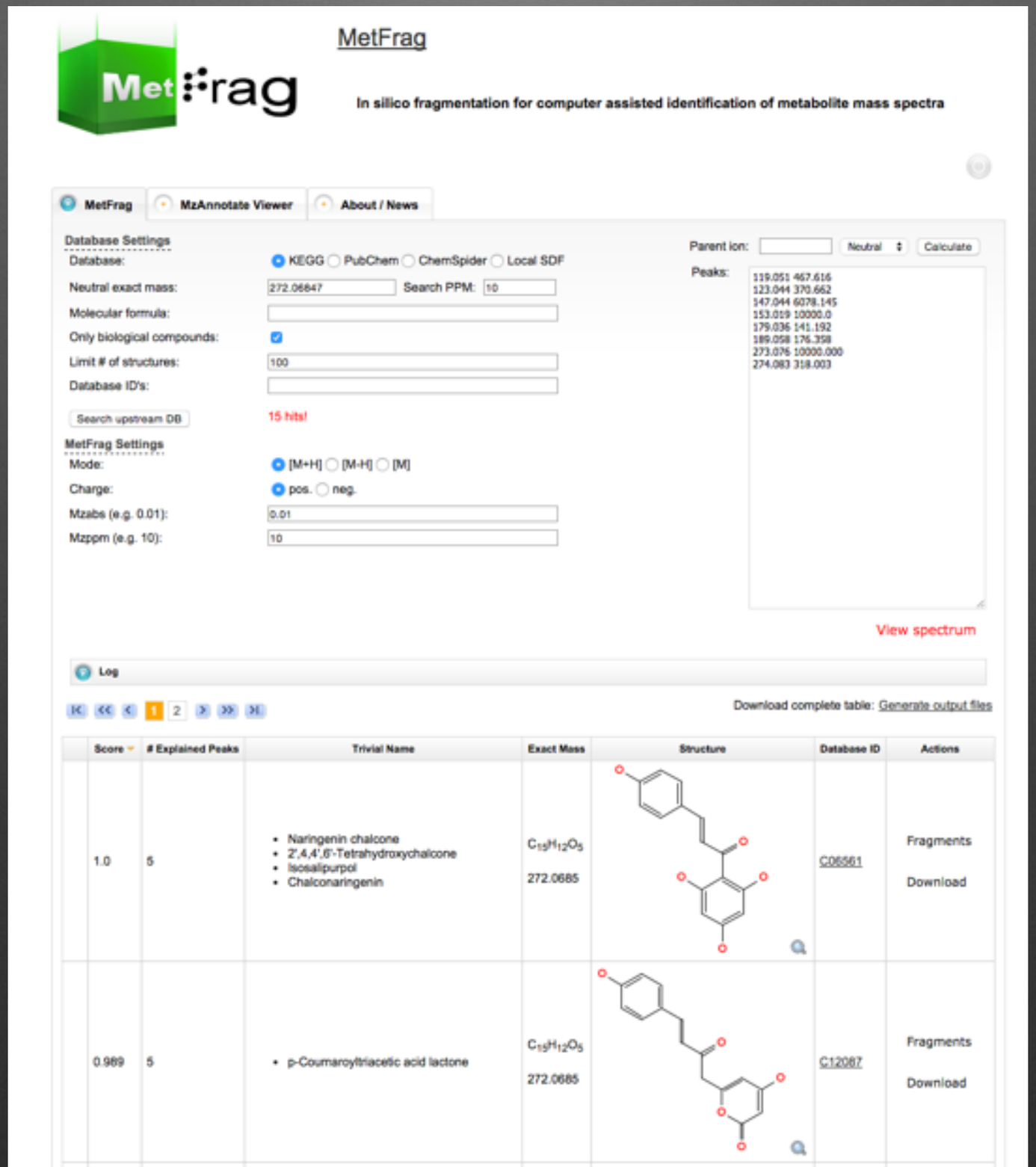
Mass-MetaSite

Agilent Technologies

Mass Frontier/
MetWorks

Thermo
SCIENTIFIC

MetFrag


MetFrag
In silico fragmentation for computer assisted identification of metabolite mass spectra
MetFrag
MzAnnotate Viewer
About / News
Database Settings
Database: KEGG PubChem ChemSpider Local SDF
Neutral exact mass: 272.06847 Search PPM: 10
Molecular formula:
Only biological compounds:
Limit # of structures: 100
Database ID's:
Search upstream DB
15 hits!
MetFrag Settings
Mode: [M+H] [M-H] [M]
Charge: pos. neg.
Mzabs (e.g. 0.01): 0.01
Mzppm (e.g. 10): 10
Parent ion:
Neutral
Calculate
Peaks:
119.051 467.616
123.044 370.662
147.044 6078.145
153.019 10000.0
179.036 141.192
189.058 176.358
273.076 10000.000
274.083 318.003
View spectrum
Log
Download complete table: Generate output files
Score
Explained Peaks
Trivial Name
Exact Mass
Structure
Database ID
Actions
1.0
5
Naringenin chalcone
2',4,4',6'-Tetrahydroxychalcone
Isosalipurpol
Chalconaringenin
C15H12O5
272.0685
C06561
Fragments
Download
0.989
5
p-Coumaroyltriacetic acid lactone
C15H12O5
272.0685
C12087
Fragments
Download

MIDAS

A DATABASE-SEARCHING ALGORITHM FOR METABOLITE IDENTIFICATION IN METABOLOMICS

NEW JOB **SUBMISSION**

Job Submitted
Job Id **8bdf8437-50fd-4ea3-816d-94e62a8e651f**

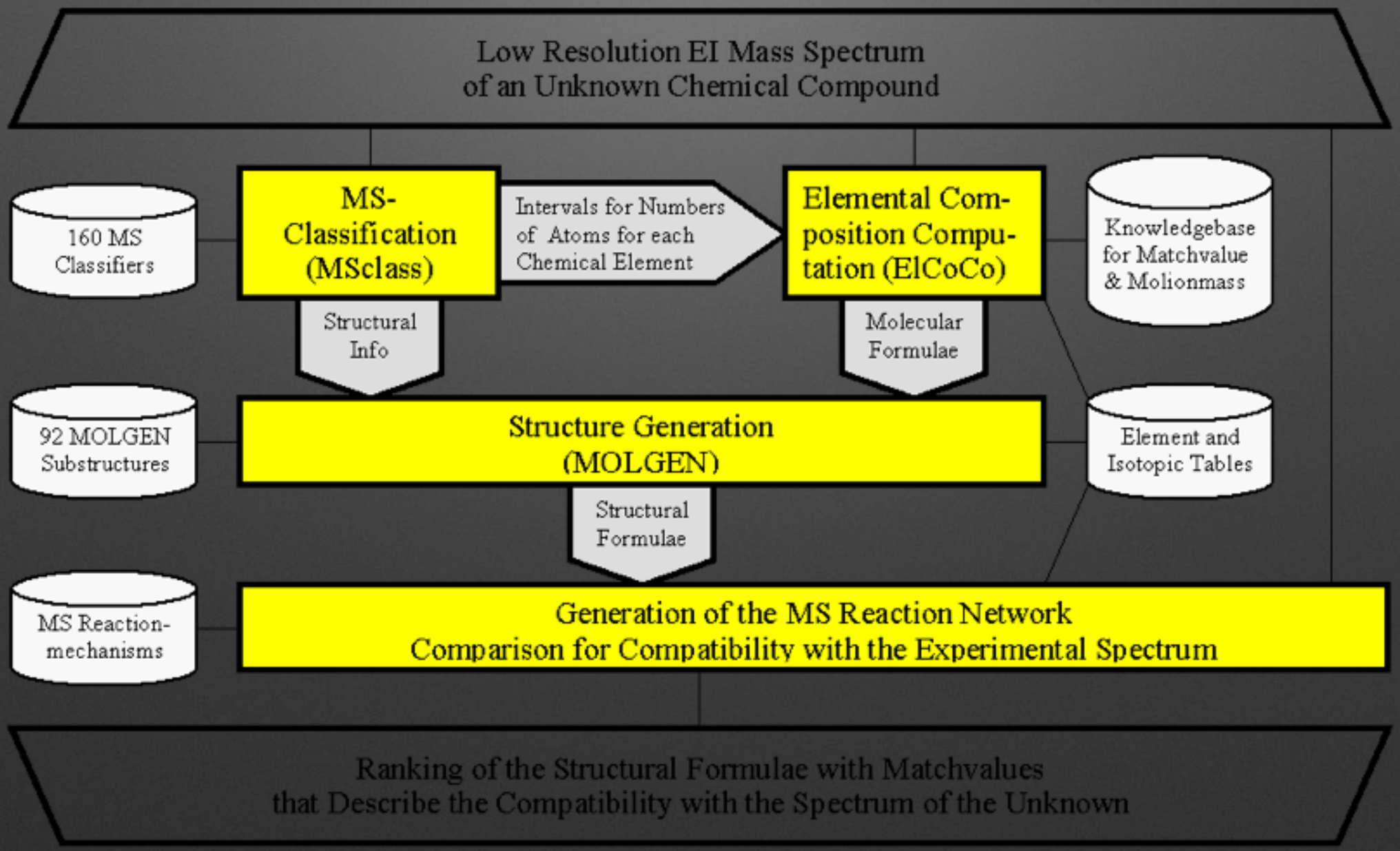
Email sent to hpbenton@gmail.com with URL to track progress.

CURRENT STATUS

☑ Job Submitted → ☐ Job In Queue → ☐ Job In Execution → ☐ Job Execution Complete

Copyright © ORNL 2014

MolGen - EI spectra

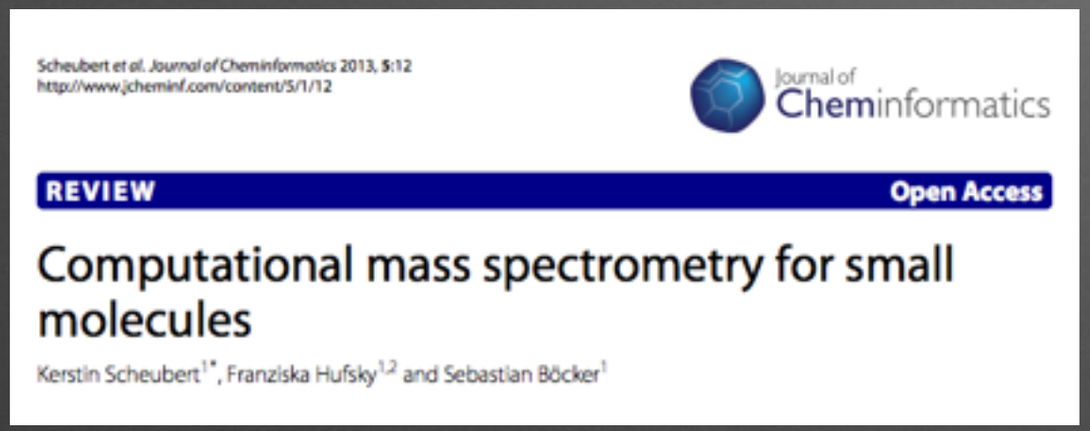


Personalised Databases

- Crowd sourcing MS/MS data
  - mzCloud - run by thermo collection of user based fragmentation data
  - Metlin - Using XCMS Online each time someone searches MS/MS data the spectra is kept in their space. If approved by user as compound X, next time search on user's personal MS/MS database.

Thank you! Questions?

Suggested reading material for more !!



Scheubert et al. Journal of Cheminformatics 2013, 5:12
http://www.jcheminf.com/content/5/1/12

Journal of Cheminformatics

REVIEW Open Access

Computational mass spectrometry for small molecules

Kerstin Scheubert[1*], Franziska Hufsky[1,2] and Sebastian Böcker[1]